PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : **C12N 15/31, C07K 14/22, 16/12, A61K 39/095 // (C12N 15/31, C12R 1:36)**	A2	(11) International Publication Number: **WO 99/24578** (43) International Publication Date: 20 May 1999 (20.05.99)

(21) International Application Number: PCT/IB98/01665

(22) International Filing Date: 9 October 1998 (09.10.98)

(30) Priority Data:

9723516.2	6 November 1997 (06.11.97)	GB
9724190.5	14 November 1997 (14.11.97)	GB
9724386.9	18 November 1997 (18.11.97)	GB
9725158.1	27 November 1997 (27.11.97)	GB
9726147.3	10 December 1997 (10.12.97)	GB
9800759.4	14 January 1998 (14.01.98)	GB
9819016.8	1 September 1998 (01.09.98)	GB

(71) Applicant *(for all designated States except US)*: CHIRON S.P.A. [IT/IT]; Via Fiorentina, 1, I-53100 Siena (IT).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: MASIGNANI, Vega [IT/IT]; Via Pantaneto, 105, I-53100 Siena (IT). RAPPUOLI, Rino [IT/IT]; Via delle Rocche, 1, Vagliagli, I-53019 Castelnuovo Berardenga (IT). PIZZA, Mariagrazia [IT/IT]; Strada di Montalbuccio, 160, I-53100 Siena (IT). SCARLATO, Vincenzo [IT/IT]; Via Firenze, 3/37, I-53134 Colle Val d'Elsa (IT). GRANDI, Guido [IT/IT]; 9° Strada, 4, I-20090 Segrate (IT).

(74) Agent: HALLYBONE, Huw, George; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: NEISSERIAL ANTIGENS

(57) Abstract

The invention provides proteins from *Neisseria meningitidis* (strains A and B) and from *Neisseria gonorrhoeae* including amino acid sequences, the corresponding nucleotide sequences, expression data, and serological data. The proteins are useful antigens for vaccines, immunogenic compositions, and/or diagnostics.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

NEISSERIAL ANTIGENS

This invention relates to antigens from *Neisseria* bacteria.

BACKGROUND ART

Neisseria meningitidis and *Neisseria gonorrhoeae* are non-motile, gram negative diplococci that are pathogenic in humans. *N.meningitidis* colonises the pharynx and causes meningitis (and, occasionally, septicaemia in the absence of meningitis); *N.gonorrhoeae* colonises the genital tract and causes gonorrhea. Although colonising different areas of the body and causing completely different diseases, the two pathogens are closely related, although one feature that clearly differentiates meningococcus from gonococcus is the presence of a polysaccharide capsule that is present in all pathogenic meningococci.

N.gonorrhoeae caused approximately 800,000 cases per year during the period 1983-1990 in the United States alone (chapter by Meitzner & Cohen, "Vaccines Against Gonococcal Infection", In: *New Generation Vaccines*, 2nd edition, ed. Levine, Woodrow, Kaper, & Cobon, Marcel Dekker, New York, 1997, pp.817-842). The disease causes significant morbidity but limited mortality. Vaccination against *N.gonorrhoeae* would be highly desirable, but repeated attempts have failed. The main candidate antigens for this vaccine are surface-exposed proteins such as pili, porins, opacity-associated proteins (Opas) and other surface-exposed proteins such as the Lip, Laz, IgA1 protease and transferrin-binding proteins. The lipooligosaccharide (LOS) has also been suggested as vaccine (Meitzner & Cohen, *supra*).

N.meningitidis causes both endemic and epidemic disease. In the United States the attack rate is 0.6-1 per 100,000 persons per year, and it can be much greater during outbreaks (see Lieberman *et al.* (1996) Safety and Immunogenicity of a Serogroups A/C *Neisseria meningitidis* Oligosaccharide-Protein Conjugate Vaccine in Young Children. *JAMA* 275(19):1499-1503; Schuchat *et al* (1997) Bacterial Meningitis in the United States in 1995. *N Engl J Med* 337(14):970-976). In developing countries, endemic disease rates are much higher and during epidemics incidence rates can reach 500 cases per 100,000 persons per year. Mortality is extremely high, at 10-20% in the United States, and much higher in developing countries. Following the introduction of the conjugate vaccine against *Haemophilus influenzae*, *N. meningitidis* is the major cause of bacterial meningitis at all ages in the United States (Schuchat *et al* (1997) *supra*).

Based on the organism's capsular polysaccharide, 12 serogroups of *N.meningitidis* have been identified. Group A is the pathogen most often implicated in epidemic disease in sub-Saharan Africa. Serogroups B and C are responsible for the vast majority of cases in the United States and in most developed countries. Serogroups W135 and Y are responsible for the rest of the cases in the United States and developed countries. The meningococcal vaccine currently in use is a tetravalent polysaccharide vaccine composed of serogroups A, C, Y and W135. Although efficacious in adolescents and adults, it induces a poor immune response and short duration of protection, and cannot be used in infants [*eg*. Morbidity and Mortality weekly report, Vol.46, No. RR-5 (1997)]. This is because polysaccharides are T-cell independent antigens that induce a weak immune response that cannot be boosted by repeated immunization. Following the success of the vaccination against *H.influenzae*, conjugate vaccines against serogroups A and C have been developed and are at the final stage of clinical testing (Zollinger WD "New and Improved Vaccines Against Meningococcal Disease" in: *New Generation Vaccines*, *supra*, pp. 469-488; Lieberman *et al* (1996) *supra*; Costantino *et al* (1992) Development and phase I clinical testing of a conjugate vaccine against meningococcus A and C. *Vaccine* 10:691-698).

Meningococcus B remains a problem, however. This serotype currently is responsible for approximately 50% of total meningitis in the United States, Europe, and South America. The polysaccharide approach cannot be used because the menB capsular polysaccharide is a polymer of α(2-8)-linked *N*-acetyl neuraminic acid that is also present in mammalian tissue. This results in tolerance to the antigen; indeed, if an immune response were elicited, it would be anti-self, and therefore undesirable. In order to avoid induction of autoimmunity and to induce a protective immune response, the capsular polysaccharide has, for instance, been chemically modified substituting the *N*-acetyl groups with N-propionyl groups, leaving the specific antigenicity unaltered (Romero & Outschoorn (1994) Current status of Meningococcal group B vaccine candidates: capsular or non-capsular? *Clin Microbiol Rev* 7(4):559-575).

Alternative approaches to menB vaccines have used complex mixtures of outer membrane proteins (OMPs), containing either the OMPs alone, or OMPs enriched in porins, or deleted of the class 4 OMPs that are believed to induce antibodies that block bactericidal activity. This approach produces vaccines that are not well characterized. They are able to protect against the homologous strain, but are not effective at large where there are many antigenic variants of the outer membrane proteins. To overcome the antigenic variability, multivalent vaccines containing up to nine different

porins have been constructed (*eg.* Poolman JT (1992) Development of a meningococcal vaccine. *Infect. Agents Dis.* 4:13-28). Additional proteins to be used in outer membrane vaccines have been the opa and opc proteins, but none of these approaches have been able to overcome the antigenic variability (*eg.* Ala'Aldeen & Borriello (1996) The meningococcal transferrin-binding proteins 1 and 2 are both surface exposed and generate bactericidal antibodies capable of killing homologous and heterologous strains. *Vaccine* 14(1):49-53).

A certain amount of sequence data is available for meningococcal and gonoccocal genes and proteins (*eg.* EP-A-0467714, WO96/29412), but this is by no means complete. The provision of further sequences could provide an opportunity to identify secreted or surface-exposed proteins that are presumed targets for the immune system and which are not antigenically variable. For instance, some of the identified proteins could be components of efficacious vaccines against meningococcus B, some could be components of vaccines against all meningococcal serotypes, and others could be components of vaccines against all pathogenic *Neisseriae*.

THE INVENTION

The invention provides proteins comprising the Neisserial amino acid sequences disclosed in the examples. These sequences relate to *N.meningitidis* or *N.gonorrhoeae*.

It also provides proteins comprising sequences homologous (*ie.* having sequence identity) to the Neisserial amino acid sequences disclosed in the examples. Depending on the particular sequence, the degree of identity is preferably greater than 50% (*eg.* 65%, 80%, 90%, or more). These homologous proteins include mutants and allelic variants of the sequences disclosed in the examples. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between the proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty=12* and *gap extension penalty=1*.

The invention further provides proteins comprising fragments of the Neisserial amino acid sequences disclosed in the examples. The fragments should comprise at least *n* consecutive amino acids from the sequences and, depending on the particular sequence, *n* is 7 or more (*eg.* 8, 10, 12, 14, 16, 18, 20 or more). Preferably the fragments comprise an epitope from the sequence.

The proteins of the invention can, of course, be prepared by various means (*eg.* recombinant expression, purification from cell culture, chemical synthesis *etc.*) and in various forms (*eg.* native, fusions *etc.*). They are preferably prepared in substantially pure or isolated form (*ie.* substantially free from other Neisserial or host cell proteins)

According to a further aspect, the invention provides antibodies which bind to these proteins. These may be polyclonal or monoclonal and may be produced by any suitable means.

According to a further aspect, the invention provides nucleic acid comprising the Neisserial nucleotide sequences disclosed in the examples. In addition, the invention provides nucleic acid comprising sequences homologous (*ie.* having sequence identity) to the Neisserial nucleotide sequences disclosed in the examples.

Furthermore, the invention provides nucleic acid which can hybridise to the Neisserial nucleic acid disclosed in the examples, preferably under "high stringency" conditions (*eg.* 65°C in a 0.1xSSC, 0.5% SDS solution).

Nucleic acid comprising fragments of these sequences are also provided. These should comprise at least *n* consecutive nucleotides from the Neisserial sequences and, depending on the particular sequence, *n* is 10 or more (*eg* 12, 14, 15, 18, 20, 25, 30, 35, 40 or more).

According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (*eg.* for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (*eg.* by chemical synthesis, from genomic or cDNA libraries, from the organism itself *etc.*) and can take various forms (*eg.* single stranded, double stranded, vectors, probes *etc.*).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (*eg.* expression vectors) and host cells transformed with such vectors.

According to a further aspect, the invention provides compositions comprising protein, antibody, and/or nucleic acid according to the invention. These compositions may be suitable as vaccines, for instance, or as diagnostic reagents, or as immunogenic compositions.

The invention also provides nucleic acid, protein, or antibody according to the invention for use as medicaments (*eg.* as vaccines) or as diagnostic reagents. It also provides the use of nucleic acid, protein, or antibody according to the invention in the manufacture of: (i) a medicament for treating or preventing infection due to Neisserial bacteria; (ii) a diagnostic reagent for detecting the presence of Neisserial bacteria or of antibodies raised against Neisserial bacteria; and/or (iii) a reagent which can raise antibodies against Neisserial bacteria. Said Neisserial bacteria may be any species or strain (such as *N.gonorrhoeae*, or any strain of *N.meningitidis*, such as strain A, strain B or strain C).

The invention also provides a method of treating a patient, comprising administering to the patient a therapeutically effective amount of nucleic acid, protein, and/or antibody according to the invention.

According to further aspects, the invention provides various processes.

A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression.

A process for producing protein or nucleic acid of the invention is provided, wherein the the protein or nucleic acid is synthesised in part or in whole using chemical means.

A process for detecting polynucleotides of the invention is provided, comprising the steps of: (a) contacting a nucleic probe according to the invention with a biological sample under hybridizing conditions to form duplexes; and (b) detecting said duplexes.

A process for detecting proteins of the invention is provided, comprising the steps of: (a) contacting an antibody according to the invention with a biological sample under conditions suitable for the formation of an antibody-antigen complexes; and (b) detecting said complexes.

A summary of standard techniques and procedures which may be employed in order to perform the invention (*eg.* to utilise the disclosed sequences for vaccination or diagnostic purposes) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

General

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature *eg.* Sambrook *Molecular Cloning; A Laboratory Manual, Second Edition* (1989); *DNA Cloning, Volumes I and ii* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I. Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

Standard abbreviations for nucleotides and amino acids are used in this specification.

All publications, patents, and patent applications cited herein are incorporated in full by reference. In particular, the contents of UK patent applications 9723516.2, 9724190.5, 9724386.9, 9725158.1, 9726147.3, 9800759.4, and 9819016.8 are incorporated herein.

Definitions

A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

The term "comprising" means "including" as well as "consisting" *eg.* a composition "comprising" X may consist exclusively of X or may include something additional to X, such as X+Y.

The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a Neisserial sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been assembled in a single protein in an arrangement not found in nature.

An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (*eg.* 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination, has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (*eg.* see US patent 5,753,235).

Expression systems

The Neisserial nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

i. Mammalian Systems

Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual, 2nd ed.]*.

Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallotheionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive cells.

The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science 236:*1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J. 4*:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al. (1982b) *Proc. Natl. Acad. Sci. 79*:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell 41*:521]. Additionally, some enhancers are regulatable and become active only

in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet. 2*:215; Maniatis et al. (1987) *Science* 236:1237].

A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus triparite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell 41*:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot (1989) *Trends Biochem. Sci. 14*:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminater/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (*eg.* plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal

viruses, which require trans-acting factors to replicate. For example, plasmids containing the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell 23*:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicaton systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol. 9*:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol. 6*:1074].

The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (*eg.* Hep G2), and a number of other cell lines.

ii. Baculovirus Systems

The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques

are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each with its owned set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (*eg.* plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) *17*:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, *42*:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli.*

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol. 69*:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, *73*:409). Alternatively, since the signals for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human α-interferon, Maeda et al., (1985), *Nature 315*:592; human gastrin-releasing peptide, Lebacq-Verheyden et al., (1988), *Molec. Cell. Biol. 8*:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, *82*:8404; mouse IL-3, (Miyajima et al., (1987) *Gene 58*:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, *7*:99, can also be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus – usually by co-transfection. The promoter

and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) *3*:2156; and Luckow and Summers (1989)). For example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays 4*:91.The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between about 1% and about 5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein, which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 μm in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus) or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers and Smith, *supra*; Miller et al. (1989).

Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti* , *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol. 56*:153; Wright (1986) *Nature 321*:718; Smith et al., (1983) *Mol. Cell. Biol. 3*:2156; and see generally, Fraser, *et al.* (1989) *In Vitro Cell. Dev. Biol. 25*:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. *See, eg.* Summers and Smith *supra*.

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as chromatography, eg. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, or the like. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, eg. proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

<u>iii. Plant Systems</u>

There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., Nucleic Acids Research 15:2515-2535 (1987); Wirsel et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R.L. Jones and J. MacMillin, Gibberellins: in: *Advanced Plant Physiology*,. Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52.

References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for Agrobacterium transformations, T DNA sequences for Agrobacterium-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilmink and Dons, 1993, *Plant Mol. Biol. Reptr*, 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet*, 202:179-185, 1985. The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, and *Datura*.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in Escherichia coli (E. coli) [Raibaud *et al.* (1984) *Annu. Rev. Genet. 18*:173]. Regulated expression may thercfore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature 198*:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res. 8*:4057; Yelverton *et al.* (1981) *Nucl. Acids Res. 9*:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The g-laotamase (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature 292*:128] and T5 [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene 25*:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci. 80*:21].

Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol. 189*:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci. 82*:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature 254*:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' and of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological Regulation and Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in Escherichia coli." In *Molecular Cloning: A Laboratory Manual*].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* on *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature 309*:810]. Fusion proteins can also be made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene 60*:197], *trpE* [Allen et al. (1987) J. Biotechnol. 5:93; Makoff *et al.*

(1989) *J. Gen. Microbiol. 135*:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology 7*:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghrayeb *et al.* (1984) *EMBO J. 3*:2437] and the *E. coli* alkaline phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci. 82*:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various Bacillus strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA 79*:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal

element (*eg.* plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various Bacillus strains integrate into the Bacillus chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol. 32*:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: Bacillus subtilis [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA 79*:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], Escherichia coli [Shimatake *et al.* (1981) *Nature 292*:128; Amann *et al.* (1985) *Gene 40*:183; Studier *et al.* (1986) *J. Mol. Biol. 189*:113; EP-A-0 036 776,EP-A-0 136 829 and EP-A-0 136 907],

Streptococcus cremoris [Powell *et al.* (1988) *Appl. Environ. Microbiol. 54*:655]; Streptococcus lividans [Powell *et al.* (1988) *Appl. Environ. Microbiol. 54*:655], Streptomyces lividans [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with $CaCl_2$ or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See *eg.* [Masson *et al.* (1989) *FEMS Microbiol. Lett. 60*:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA 79*:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, Bacillus], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci. 85*:856; Wang *et al.* (1990) *J. Bacteriol. 172*:949, Campylobacter], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci. 69*:2110; Dower *et al.* (1988) *Nucleic Acids Res. 16*:6127; Kushner (1978) "An improved method for transformation of Escherichia coli with ColE1-derived plasmids. In *Genetic Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol. 53*:159; Taketo (1988) *Biochim. Biophys. Acta 949*:318; Escherichia], [Chassy *et al.* (1987) *FEMS Microbiol. Lett. 44*:173 Lactobacillus]; [Fiedler *et al.* (1988) *Anal. Biochem 170*:38, Pseudomonas]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett. 66*:203, Staphylococcus], [Barany *et al.* (1980) *J. Bacteriol. 144*:698; Harlander (1987) "Transformation of Streptococcus lactis by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun. 32*:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol. 54*:655; Somkuti *et al.* (1987) *Proc. 4th Evr. Cong. Biotechnology 1*:412, Streptococcus].

v. Yeast Expression

Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (*eg.* structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence

of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA 80*:1].

In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*, *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA 77*:1078; Henikoff *et al.* (1981) *Nature 283*:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol. 96*:119; Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast Saccharomyces cerevisiae," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene 11*:163; Panthier *et al.* (1980) *Curr. Genet. 2*:109;].

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *eg.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*eg.* ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*eg.* WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the yeast invertase gene (EP-A-0 012 873; JPO. 62,096,086) and the A-factor gene (US patent 4,588,684). Alternatively, leaders of non-yeast origin, such as an interferon leader, exist that also provide for secretion in yeast (EP-A-0 060 057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (*eg.* see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (*eg.* plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein *et al.* (1979) *Gene 8*:17-24], pCl/1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA 81*:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol. 158*:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See *eg.* Brake *et al.*, *supra.*

Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol. 101*:228-245]. An integrating vector may be directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra.* One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA 80*:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the

chromosome and flanking the expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol, Rev. 51*:351].

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts:Candida albicans [Kurtz, *et al.* (1986) *Mol. Cell. Biol. 6*:142], Candida maltosa [Kunze, *et al.* (1985) *J. Basic Microbiol. 25*:141]. Hansenula polymorpha [Gleeson, *et al.* (1986) *J. Gen. Microbiol. 132*:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet. 202*:302], Kluyveromyces fragilis [Das, *et al.* (1984) *J. Bacteriol. 158*:1165], Kluyveromyces lactis [De Louvencourt *et al.* (1983) *J. Bacteriol. 154*:737; Van den Berg *et al.* (1990) *Bio/Technology 8*:135], Pichia guillerimondii [Kunze *et al.* (1985) *J. Basic Microbiol. 25*:141], Pichia pastoris [Cregg, *et al.* (1985) *Mol. Cell. Biol. 5*:3376; US Patent Nos. 4,837,148 and 4,929,555], Saccharomyces cerevisiae [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA 75*:1929; Ito *et al.* (1983) *J. Bacteriol. 153*:163], Schizosaccharomyces pombe [Beach and Nurse (1981) *Nature 300*:706], and Yarrowia lipolytica [Davidow, *et al.* (1985) *Curr. Genet. 10*:380471 Gaillardin, *et al.* (1985) *Curr. Genet. 10*:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See *eg.* [Kurtz *et al.* (1986) *Mol. Cell. Biol. 6*:142; Kunze *et al.* (1985) *J. Basic Microbiol. 25*:141; Candida];

[Gleeson *et al.* (1986) *J. Gen. Microbiol. 132*:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet. 202*:302; Hansenula]; [Das *et al.* (1984) *J. Bacteriol. 158*:1165; De Louvencourt *et al.* (1983) *J. Bacteriol. 154*:1165; Van den Berg *et al.* (1990) *Bio/Technology 8*:135; Kluyveromyces]; [Cregg *et al.* (1985) *Mol. Cell. Biol. 5*:3376; Kunze *et al.* (1985) *J. Basic Microbiol. 25*:141; US Patent Nos. 4,837,148 and 4,929,555; Pichia]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA 75*;1929; Ito *et al.* (1983) *J. Bacteriol. 153*:163 Saccharomyces]; [Beach and Nurse (1981) *Nature 300*:706; Schizosaccharomyces]; [Davidow *et al.* (1985) *Curr. Genet. 10*:39; Gaillardin *et al.* (1985) *Curr. Genet. 10*:49; Yarrowia].

Antibodies

As used herein, the term "antibody" refers to a polypeptide or group of polypeptides composed of at least one antibody combining site. An "antibody combining site" is the three-dimensional binding space with an internal surface shape and charge distribution complementary to the features of an epitope of an antigen, which allows a binding of the antibody with the antigen. "Antibody" includes, for example, vertebrate antibodies, hybrid antibodies, chimeric antibodies, humanised antibodies, altered antibodies, univalent antibodies, Fab proteins, and single domain antibodies.

Antibodies against the proteins of the invention are useful for affinity chromatography, immunoassays, and distinguishing/identifying Neisserial proteins.

Antibodies to the proteins of the invention, both polyclonal and monoclonal, may be prepared by conventional methods. In general, the protein is first used to immunize a suitable animal, preferably a mouse, rat, rabbit or goat. Rabbits and goats are preferred for the preparation of polyclonal sera due to the volume of serum obtainable, and the availability of labeled anti-rabbit and anti-goat antibodies. Immunization is generally performed by mixing or emulsifying the protein in saline, preferably in an adjuvant such as Freund's complete adjuvant, and injecting the mixture or emulsion parenterally (generally subcutaneously or intramuscularly). A dose of 50-200 μg/injection is typically sufficient. Immunization is generally boosted 2-6 weeks later with one or more injections of the protein in saline, preferably using Freund's incomplete adjuvant. One may alternatively generate antibodies by in vitro immunization using methods known in the art, which for the purposes of this invention is considered equivalent to *in vivo* immunization. Polyclonal antisera is obtained by bleeding the immunized animal into a glass or plastic container, incubating the blood at 25°C for one hour, followed by incubating at 4°C for 2-18 hours. The serum is

recovered by centrifugation (*eg.* 1,000*g* for 10 minutes). About 20-50 ml per bleed may be obtained from rabbits.

Monoclonal antibodies are prepared using the standard method of Kohler & Milstein [*Nature* (1975) 256:495-96], or a modification thereof. Typically, a mouse or rat is immunized as described above. However, rather than bleeding the animal to extract serum, the spleen (and optionally several large lymph nodes) is removed and dissociated into single cells. If desired, the spleen cells may be screened (after removal of nonspecifically adherent cells) by applying a cell suspension to a plate or well coated with the protein antigen. B-cells expressing membrane-bound immunoglobulin specific for the antigen bind to the plate, and are not rinsed away with the rest of the suspension. Resulting B-cells, or all dissociated spleen cells, are then induced to fuse with myeloma cells to form hybridomas, and are cultured in a selective medium (*eg.* hypoxanthine, aminopterin, thymidine medium, "HAT"). The resulting hybridomas are plated by limiting dilution, and are assayed for the production of antibodies which bind specifically to the immunizing antigen (and which do not bind to unrelated antigens). The selected MAb-secreting hybridomas are then cultured either *in vitro* (*eg.* in tissue culture bottles or hollow fiber reactors), or *in vivo* (as ascites in mice).

If desired, the antibodies (whether polyclonal or monoclonal) may be labeled using conventional techniques. Suitable labels include fluorophores, chromophores, radioactive atoms (particularly ^{32}P and ^{125}I), electron-dense reagents, enzymes, and ligands having specific binding partners. Enzymes are typically detected by their activity. For example, horseradish peroxidase is usually detected by its ability to convert 3,3',5,5'-tetramethylbenzidine (TMB) to a blue pigment, quantifiable with a spectrophotometer. "Specific binding partner" refers to a protein capable of binding a ligand molecule with high specificity, as for example in the case of an antigen and a monoclonal antibody specific therefor. Other specific binding partners include biotin and avidin or streptavidin, IgG and protein A, and the numerous receptor-ligand couples known in the art. It should be understood that the above description is not meant to categorize the various labels into distinct classes, as the same label may serve in several different modes. For example, ^{125}I may serve as a radioactive label or as an electron-dense reagent. HRP may serve as enzyme or as antigen for a MAb. Further, one may combine various labels for desired effect. For example, MAbs and avidin also require labels in the practice of this invention: thus, one might label a MAb with biotin, and detect its presence with avidin labeled with ^{125}I, or with an anti-biotin MAb labeled with HRP. Other permutations and possibilities will be

readily apparent to those of ordinary skill in the art, and are considered as equivalents within the scope of the instant invention.

Pharmaceutical Compositions

Pharmaceutical compositions can comprise either polypeptides, antibodies, or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (*eg.* see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Vaccines

Vaccines according to the invention may either be prophylactic (*ie.* to prevent infection) or therapeutic (*ie.* to treat disease after infection).

Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, *etc.* pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents

such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in *Vaccine design: the subunit and adjuvant approach*, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (*eg.* IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, *etc.*), interferons (*eg.* gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59™ are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-*sn*-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), *etc.*

The immunogenic compositions (*eg.* the immunising antigen/immunogen/polypeptide/protein/ nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (*eg.* nonhuman primate, primate, *etc.*), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, *eg.* by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (*eg.* WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed [*eg.* Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; see later herein].

Gene Delivery Vehicles

Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence in vivo can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus,

picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses *eg.* MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, Second Edition, Cold Spring Harbor Laboratory, 1985.

Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (*eg.* HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia, Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC Nol VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or

collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671, WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the

native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar Institute), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 and WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those deposited with the ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995,WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

DNA vector systems such as eukarytic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems. Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, Nature 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805; Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Triniti virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid

expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA*

91(24):11581-11585. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; inWO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, Biochemistry, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.

A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

Delivery Methods

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for expression of recombinant proteins. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (*eg.* see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in *eg.* WO93/14778. Examples of cells useful in ex vivo applications include, for example, stem cells, particularly hematopoetic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A.Polypeptides

One example are polypeptides which include, without limitation: asioloorosomucoid (ASOR); transferrin; asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

B.Hormones, Vitamins, *etc.*

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C.Polyalkylenes, Polysaccharides, *etc.*

Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethlylene glycol. In addition, mono-, di-, or polysaccarides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D.Lipids, and Liposomes

The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the

use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta.* 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See, also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boerhinger). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, *eg.* Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphoshatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilammelar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See *eg.* Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta* 394:483; Wilson (1979) *Cell* 17:77); Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348); Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

E.Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C, and E, over time these lipoproteins lose A and acquire C and E apoproteins. VLDL comprises A, B, C, and E apoproteins, LDL comprises apoprotein B; and HDL comprises apoproteins A, C, and E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) Annu Rev. Biochem 54:699; Law (1986) Adv. Exp Med. Biol. 151:162; Chen (1986) J Biol Chem 261:12918; Kane (1980) Proc Natl Acad Sci USA 77:2465; and Utermann (1984) Hum Genet 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phopholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol.* (*supra*); Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30:

443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Techniologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

F.Polycationic Agents

Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both in vitro, ex vivo, and in vivo applications. Polycationic agents can be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, etc.

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful as nucleic aid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and purtrescine.

The dimensions and of the physical properties of a polycationic agent can be extrapolated from the list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when combined with polynucleotides/polypeptides.

Immunodiagnostic Assays

Neisserial antigens of the invention can be used in immunoassays to detect antibody levels (or, conversely, anti-Neisserial antibodies can be used to detect antigen levels). Immunoassays based on well defined, recombinant antigens can be developed to replace invasive diagnostics methods. Antibodies to Neisserial proteins within biological samples, including for example, blood or serum

samples, can be detected. Design of the immunoassays is subject to a great deal of variation, and a variety of these are known in the art. Protocols for the immunoassay may be based, for example, upon competition, or direct reaction, or sandwich type assays. Protocols may also, for example, use solid supports, or may be by immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the probe are also known; examples of which are assays which utilize biotin and avidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

Kits suitable for immunodiagnosis and containing the appropriate labeled reagents are constructed by packaging the appropriate materials, including the compositions of the invention, in suitable containers, along with the remaining reagents and materials (for example, suitable buffers, salt solutions, *etc.*) required for the conduct of the assay, as well as suitable set of assay instructions.

Nucleic Acid Hybridisation

"Hybridization" refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [*supra*] Volume 2, chapter 9, pages 9.47 to 9.57.

"Stringency" refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated Tm of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The

total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1μg for a plasmid or phage digest to 10^{-9} to 10^{-8} g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 μg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10^8 cpm/μg. For a single-copy mammalian gene a conservative approach would start with 10 μg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10^8 cpm/μg, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature (Tm) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$Tm= 81 + 16.6(\log_{10}Ci) + 0.4[\%(G + C)]-0.6(\%formamide) - 600/n-1.5(\%mismatch).$$

where Ci is the salt concentration (monovalent ions) and *n* is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with is 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology,

and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

The nucleic acid probes will hybridize to the Neisserial nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native Neisserial sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the Neisserial sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Neisserial sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a Neisserial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a Neisserial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as temperature, salt condition and the like. For example, for diagnostic applications, depending on the

complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably at least 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *eg.* backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*eg.* see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as peptide nucleic acids may also be used [*eg.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].

Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* [*Meth. Enzymol.* (1987) 155: 335-350]; US patents 4,683,195 and 4,683,202. Two "primer" nucleotides hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its complement) to aid with duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired Neisserial sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern blots. When using the Southern blot method, the labelled probe will hybridize to the Neisserial sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed

to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1-20 show biochemical data obtained in the Examples, and also sequence analysis, for ORFs 37, 5, 2, 15, 22, 28, 32, 4, 61, 76, 89, 97, 106, 138, 23, 25, 27, 79, 85 and 132. M1 and M2 are molecular weight markers. Arrows indicate the position of the main recombinant product or, in Western blots, the position of the main *N.meningitidis* immunoreactive band. TP indicates *N.meningitidis* total protein extract; OMV indicates *N.meningitidis* outer membrane vesicle preparation. In bactericidal assay results: a diamond (♦) shows preimmune data; a triangle (▲) shows GST control data; a circle (●) shows data with recombinant *N.meningitidis* protein. Computer analyses show a hydrophilicity plot (upper), an antigenic index plot (middle), and an AMPHI analysis (lower). The AMPHI program has been used to predict T-cell epitopes [Gao *et al.* (1989) *J. Immunol.* 143:3007; Roberts *et al.* (1996) *AIDS Res Hum Retrovir* 12:593; Quakyi *et al.* (1992) *Scand J Immunol* suppl.11:9) and is available in the Protean package of DNASTAR, Inc. (1228 South Park Street, Madison, Wisconsin 53715 USA).

EXAMPLES

The examples describe nucleic acid sequences which have been identified in *N.meningitidis*, along with their putative translation products, and also those of *N.gonorrhoeae*. Not all of the nucleic acid sequences are complete *ie.* they encode less than the full-length wild-type protein.

The examples are generally in the following format:

- a nucleotide sequence which has been identified in *N.meningitidis* (strain B)
- the putative translation product of this sequence
- a computer analysis of the translation product based on database comparisons
- corresponding gene and protein sequences identified in *N.meningitidis* (strain A) and in *N.gonorrhoeae*
- a description of the characteristics of the proteins which indicates that they might be suitably antigenic
- results of biochemical analysis (expression, purification, ELISA, FACS *etc.*)

The examples typically include details of sequence identity between species and strains. Proteins that are similar in sequence are generally similar in both structure and function, and the sequence identity often indicates a common evolutionary origin. Comparison with sequences of proteins of known function is widely used as a guide for the assignment of putative protein function to a new sequence and has proved particularly useful in whole-genome analyses.

Sequence comparisons were performed at NCBI (http://www.ncbi.nlm.nih.gov) using the algorithms BLAST, BLAST2, BLASTn, BLASTp, tBLASTn, BLASTx, & tBLASTx [*eg.* see also Altschul *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:2289-3402]. Searches were performed against the following databases: non-redundant GenBank+EMBL+DDBJ+PDB sequences and non-redundant GenBank CDS translations+PDB+SwissProt+SPupdate+PIR sequences.

To compare Meningococcal and Gonococcal sequences, the tBLASTx algorithm was used, as implemented at http://www.genome.ou.edu/gono_blast.html. The FASTA algorithm was also used to compare the ORFs (from GCG Wisconsin Package, version 9.0).

Dots within nucleotide sequences (*eg.* position 495 in SEQ ID 11) represent nucleotides which have been arbitrarily introduced in order to maintain a reading frame. In the same way, double-underlined nucleotides were removed. Lower case letters (*eg.* position 496 in SEQ ID 11) represent ambiguities which arose during alignment of independent sequencing reactions (some of the nucleotide sequences in the examples are derived from combining the results of two or more experiments).

Nucleotide sequences were scanned in all six reading frames to predict the presence of hydrophobic domains using an algorithm based on the statistical studies of Esposti *et al.* [Critical evaluation of the hydropathy of membrane proteins (1990) *Eur J Biochem* 190:207-219]. These domains represent potential transmembrane regions or hydrophobic leader sequences.

Open reading frames were predicted from fragmented nucleotide sequences using the program ORFFINDER (NCBI).

Underlined amino acid sequences indicate possible transmembrane domains or leader sequences in the ORFs, as predicted by the PSORT algorithm (http://www.psort.nibb.ac.jp). Functional domains were also predicted using the MOTIFS program (GCG Wisconsin & PROSITE).

Various tests can be used to assess the *in vivo* immunogencity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question *ie.* the protein is an immunogen. This method can also be used to identify immunodominant proteins.

The recombinant protein can also be conveniently used to prepare antibodies *eg.* in a mouse. These can be used for direct confirmation that a protein is located on the cell-surface. Labelled antibody (*eg.* fluorescent labelling for FACS) can be incubated with intact bacteria and the presence of label on the bacterial surface confirms the location of the protein.

In particular, the following methods (A) to (S) were used to express, purify and biochemically characterise the proteins of the invention:

A) Chromosomal DNA preparation

N.meningitidis strain 2996 was grown to exponential phase in 100ml of GC medium, harvested by centrifugation, and resuspended in 5ml buffer (20% Sucrose, 50mM Tris-HCl, 50mM EDTA, pH8). After 10 minutes incubation on ice, the bacteria were lysed by adding 10ml lysis solution (50mM NaCl, 1% Na-Sarkosyl, 50μg/ml Proteinase K), and the suspension was incubated at 37°C for 2 hours. Two phenol extractions (equilibrated to pH 8) and one $ChCl_3$/isoamylalcohol (24:1) extraction were performed. DNA was precipitated by addition of 0.3M sodium acetate and 2 volumes ethanol, and was collected by centrifugation. The pellet was washed once with 70% ethanol and redissolved in 4ml buffer (10mM Tris-HCl, 1mM EDTA, pH 8). The DNA concentration was measured by reading the OD at 260 nm.

B) Oligonucleotide design

Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF, using (a) the meningococcus B sequence when available, or (b) the gonococcus/meningococcus A sequence, adapted to the codon preference usage of meningococcus as necessary. Any predicted signal peptides were omitted, by deducing the 5'-end amplification primer sequence immediately downstream from the predicted leader sequence.

For most ORFs, the 5' primers included two restriction enzyme recognition sites (*BamH*I-*Nde*I, *BamH*I-*Nhe*I, or *EcoR*I-*Nhe*I, depending on the gene's own restriction pattern); the 3' primers included

a *Xho*I restriction site. This procedure was established in order to direct the cloning of each amplification product (corresponding to each ORF) into two different expression systems: pGEX-KG (using either *BamH*I-*Xho*I or *EcoR*I-*Xho*I), and pET21b+ (using either *Nde*I-*Xho*I or *Nhe*I-*Xho*I).

5'-end primer tail:	CGCGGATCCCATATG	(*BamH*I-*Nde*I)
	CGCGGATCCGCTAGC	(*BamH*I-*Nhe*I)
	CCGGAATTCTAGCTAGC	(*EcoR*I-*Nhe*I)
3'-end primer tail:	CCCGCTCGAG	(*Xho*I)

For ORFs 5, 15, 17, 19, 20, 22, 27, 28, 65 & 89, two different amplifications were performed to clone each ORF in the two expression systems. Two different 5' primers were used for each ORF; the same 3' *Xho*I primer was used as before:

5'-end primer tail:	GGAATTCCATATGGCCATGG	(*Nde*I)
5'-end primer tail:	CGGGATCC	(*BamH*I)

ORF 76 was cloned in the pTRC expression vector and expressed as an amino-terminus His-tag fusion. In this particular case, the predicted signal peptide was included in the final product. *Nhe*I-*BamH*I restriction sites were incorporated using primers:

5'-end primer tail:	GATCAGCTAGCCATATG	(*Nhe*I)
3'-end primer tail:	CGGGATCC	(*BamH*I)

As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridizeed to the sequence to be amplified. The number of hybridizing nucleotides depended on the melting temperature of the whole primer, and was determined for each primer using the formulae:

$$T_m = 4\ (G+C) + 2\ (A+T) \qquad \text{(tail excluded)}$$

$$T_m = 64.9 + 0.41\ (\%\ GC) - 600/N \qquad \text{(whole primer)}$$

The average melting temperature of the selected oligos were 65-70°C for the whole oligo and 50-55°C for the hybridising region alone.

Table I (page 487) shows the forward and reverse primers used for each amplification. In certain cases, it will be noted that the sequence of the primer does not exactly match the sequence in the ORF. When initial amplifications were performed, the complete 5' and/or 3' sequence was not

known for some meningococcal ORFs, although the corresponding sequences had been identified in gonococcus. For amplification, the gonococcal sequences could thus be used as the basis for primer design, altered to take account of codon preference. In particular, the following codons were changed: ATA→ATT; TCG→TCT; CAG→CAA; AAG→AAA; GAG→GAA; CGA→CGC; CGG→CGC; GGG→GGC. Italicised nucleotides in Table I indicate such a change. It will be appreciated that, once the complete sequence has been identified, this approach is generally no longer necessary.

Oligos were synthesized by a Perkin Elmer 394 DNA/RNA Synthesizer, eluted from the columns in 2ml NH_4OH, and deprotected by 5 hours incubation at 56°C. The oligos were precipitated by addition of 0.3M Na-Acetate and 2 volumes ethanol. The samples were then centrifuged and the pellets resuspended in either 100μl or 1ml of water. OD_{260} was determined using a Perkin Elmer Lambda Bio spectophotometer and the concentration was determined and adjusted to 2-10pmol/μl.

C) Amplification

The standard PCR protocol was as follows: 50-200ng of genomic DNA were used as a template in the presence of 20-40μM of each oligo, 400-800μM dNTPs solution, 1x PCR buffer (including 1.5mM $MgCl_2$), 2.5 units *TaqI* DNA polymerase (using Perkin-Elmer AmpliTaQ, GIBCO Platinum, Pwo DNA polymerase, or Tahara Shuzo Taq polymerase).

In some cases, PCR was optimsed by the addition of 10μl DMSO or 50μl 2M betaine.

After a hot start (adding the polymerase during a preliminary 3 minute incubation of the whole mix at 95°C), each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridization temperature the one of the oligos excluding the restriction enzymes tail, followed by 30 cycles performed according to the hybridization temperature of the whole length oligos. The cycles were followed by a final 10 minute extension step at 72°C.

The standard cycles were as follows:

	Denaturation	**Hybridisation**	**Elongation**
First 5 cycles	30 seconds 95°C	30 seconds 50-55°C	30-60 seconds 72°C
Last 30 cycles	30 seconds	30 seconds	30-60 seconds

	95°C	65-70°C	72°C

The elongation time varied according to the length of the ORF to be amplified.

The amplifications were performed using either a 9600 or a 2400 Perkin Elmer GeneAmp PCR System. To check the results, 1/10 of the amplification volume was loaded onto a 1-1.5% agarose gel and the size of each amplified fragment compared with a DNA molecular weight marker.

The amplified DNA was either loaded directly on a 1% agarose gel or first precipitated with ethanol and resuspended in a suitable volume to be loaded on a 1% agarose gel. The DNA fragment corresponding to the right size band was then eluted and purified from gel, using the Qiagen Gel Extraction Kit, following the instructions of the manufacturer. The final volume of the DNA fragment was 30μl or 50μl of either water or 10mM Tris, pH 8.5.

D) Digestion of PCR fragments

The purified DNA corresponding to the amplified fragment was split into 2 aliquots and double-digested with:

- *NdeI/XhoI* or *NheI/XhoI* for cloning into pET-21b+ and further expression of the protein as a C-terminus His-tag fusion
- *BamHI/XhoI* or *EcoRI/XhoI* for cloning into pGEX-KG and further expression of the protein as N-terminus GST fusion.
- For ORF 76, *NheI/BamHI* for cloning into pTRC-HisA vector and further expression of the protein as N-terminus His-tag fusion.
- *EcoRI/PstI, EcoRI/SalI, SalI/PstI* for cloning into pGex-His and further expression of the protein as N-terminus His-tag fusion

Each purified DNA fragment was incubated (37°C for 3 hours to overnight) with 20 units of each restriction enzyme (New England Biolabs) in a either 30 or 40μl final volume in the presence of the appropriate buffer. The digestion product was then purified using the QIAquick PCR purification kit, following the manufacturer's instructions, and eluted in a final volume of 30 or 50μl of either water or 10mM Tris-HCl, pH 8.5. The final DNA concentration was determined by 1% agarose gel electrophoresis in the presence of titrated molecular weight marker.

E) Digestion of the cloning vectors (pET22B, pGEX-KG, pTRC-His A, and pGex-His)

10μg plasmid was double-digested with 50 units of each restriction enzyme in 200μl reaction volume in the presence of appropriate buffer by overnight incubation at 37°C. After loading the whole digestion on a 1% agarose gel, the band corresponding to the digested vector was purified from the gel using the Qiagen QIAquick Gel Extraction Kit and the DNA was eluted in 50μl of 10mM Tris-HCl, pH 8.5. The DNA concentration was evaluated by measuring OD_{260} of the sample, and adjusted to 50μg/μl. 1μl of plasmid was used for each cloning procedure.

The vector pGEX-His is a modified pGEX-2T vector carrying a region encoding six histidine residues upstream to the thrombin cleavage site and containing the multiple cloning site of the vector pTRC99 (Pharmacia).

F) Cloning

The fragments corresponding to each ORF, previously digested and purified, were ligated in both pET22b and pGEX-KG. In a final volume of 20μl, a molar ratio of 3:1 fragment/vector was ligated using 0.5μl of NEB T4 DNA ligase (400 units/μl), in the presence of the buffer supplied by the manufacturer. The reaction was incubated at room temperature for 3 hours. In some experiments, ligation was performed using the Boheringer "Rapid Ligation Kit", following the manufacturer's instructions.

In order to introduce the recombinant plasmid in a suitable strain, 100μl *E. coli* DH5 competent cells were incubated with the ligase reaction solution for 40 minutes on ice, then at 37°C for 3 minutes, then, after adding 800μl LB broth, again at 37°C for 20 minutes. The cells were then centrifuged at maximum speed in an Eppendorf microfuge and resuspended in approximately 200μl of the supernatant. The suspension was then plated on LB ampicillin (100mg/ml).

The screening of the recombinant clones was performed by growing 5 randomly-chosen colonies overnight at 37°C in either 2ml (pGEX or pTC clones) or 5ml (pET clones) LB broth + 100μg/ml ampicillin. The cells were then pelletted and the DNA extracted using the Qiagen QIAprep Spin Miniprep Kit, following the manufacturer's instructions, to a final volume of 30μl. 5μl of each individual miniprep (approximately 1g) were digested with either *NdeI/XhoI* or *BamHI/XhoI* and the whole digestion loaded onto a 1-1.5% agarose gel (depending on the expected insert size), in parallel with the molecular weight marker (1Kb DNA Ladder, GIBCO). The screening of the positive clones was made on the base of the correct insert size.

For the cloning of ORFs 110, 111, 113, 115, 119, 122, 125 & 130, the double-digested PCR product was ligated into double-digested vector using *EcoR*I-*Pst*I cloning sites or, for ORFs 115 & 127, *EcoR*I-*Sal*I or, for ORF 122, *Sal*I-*Pst*I. After cloning, the recombinant plasmids were introduced in the *E.coli* host W3110. Individual clones were grown overnight at 37°C in L-broth with 50μl/ml ampicillin.

G) Expression

Each ORF cloned into the expression vector was transformed into the strain suitable for expression of the recombinant protein product. 1μl of each construct was used to transform 30μl of *E.coli* BL21 (pGEX vector), *E.coli* TOP 10 (pTRC vector) or *E.coli* BL21-DE3 (pET vector), as described above. In the case of the pGEX-His vector, the same *E.coli* strain (W3110) was used for initial cloning and expression. Single recombinant colonies were inoculated into 2ml LB+Amp (100μg/ml), incubated at 37°C overnight, then diluted 1:30 in 20ml of LB+Amp (100μg/ml) in 100ml flasks, making sure that the OD_{600} ranged between 0.1 and 0.15. The flasks were incubated at 30°C into gyratory water bath shakers until OD indicated exponential growth suitable for induction of expression (0.4-0.8 OD for pET and pTRC vectors; 0.8-1 OD for pGEX and pGEX-His vectors). For the pET, pTRC and pGEX-His vectors, the protein expression was induced by addition of 1mM IPTG, whereas in the case of pGEX system the final concentration of IPTG was 0.2mM. After 3 hours incubation at 30°C, the final concentration of the sample was checked by OD. In order to check expression, 1ml of each sample was removed, centrifuged in a microfuge, the pellet resuspended in PBS, and analysed by 12% SDS-PAGE with Coomassie Blue staining. The whole sample was centrifuged at 6000*g* and the pellet resuspended in PBS for further use.

H) GST-fusion proteins large-scale purification.

A single colony was grown overnight at 37°C on LB+Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid colture in a water bath shaker and grown overnight. Bacteria were diluted 1:30 into 600ml of fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD_{550} 0.8-1. Protein expression was induced with 0.2mM IPTG followed by three hours incubation. The culture was centrifuged at 8000rpm at 4°C. The supernatant was discarded and the bacterial pellet was resuspended in 7.5ml cold PBS. The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and centrifuged again. The supernatant was collected and mixed with 150μl Glutatione-Sepharose 4B resin (Pharmacia)

(previously washed with PBS) and incubated at room temperature for 30 minutes. The sample was centrifuged at 700*g* for 5 minutes at 4°C. The resin was washed twice with 10ml cold PBS for 10 minutes, resuspended in 1ml cold PBS, and loaded on a disposable column. The resin was washed twice with 2ml cold PBS until the flow-through reached OD_{280} of 0.02-0.06. The GST-fusion protein was eluted by addition of 700μl cold Glutathione elution buffer (10mM reduced glutathione, 50mM Tris-HCl) and fractions collected until the OD_{280} was 0.1. 21μl of each fraction were loaded on a 12% SDS gel using either Biorad SDS-PAGE Molecular weight standard broad range (M1) (200, 116.25, 97.4, 66.2, 45, 31, 21.5, 14.4, 6.5 kDa) or Amersham Rainbow Marker (M2) (220, 66, 46, 30, 21.5, 14.3 kDa) as standards. As the MW of GST is 26kDa, this value must be added to the MW of each GST-fusion protein.

I) His-fusion solubility analysis (ORFs 111-129)

To analyse the solubility of the His-fusion expression products, pellets of 3ml cultures were resuspended in buffer M1 [500μl PBS pH 7.2]. 25μl lysozyme (10mg/ml) was added and the bacteria were incubated for 15 min at 4°C. The pellets were sonicated for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed twice and then separated again into pellet and supernatant by a centrifugation step. The supernatant was collected and the pellet was resuspended in buffer M2 [8M urea, 0.5M NaCl, 20mM imidazole and 0.1M NaH_2PO_4] and incubated for 3 to 4 hours at 4°C. After centrifugation, the supernatant was collected and the pellet was resuspended in buffer M3 [6M guanidinium-HCl, 0.5M NaCl, 20mM imidazole and 0.1M NaH_2PO_4] overnight at 4°C. The supernatants from all steps were analysed by SDS-PAGE.

The proteins expressed from ORFs 113, 119 and 120 were found to be soluble in PBS, whereas ORFs 111, 122, 126 and 129 need urea and ORFs 125 and 127 need guanidium-HCl for their solubilization.

J) His-fusion large-scale purification.

A single colony was grown overnight at 37°C on a LB + Amp agar plate. The bacteria were inoculated into 20ml of LB+Amp liquid culture and incubated overnight in a water bath shaker. Bacteria were diluted 1:30 into 600ml fresh medium and allowed to grow at the optimal temperature (20-37°C) to OD_{550} 0.6-0.8. Protein expression was induced by addition of 1mM IPTG and the culture further incubated for three hours. The culture was centrifuged at 8000rpm at 4°C, the supernatant was discarded and the bacterial pellet was resuspended in 7.5ml of either (i) cold

buffer A (300mM NaCl, 50mM phosphate buffer, 10mM imidazole, pH 8) for soluble proteins or (ii) buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 8.8) for insoluble proteins.

The cells were disrupted by sonication on ice for 30 sec at 40W using a Branson sonifier B-15, frozen and thawed two times and centrifuged again.

For insoluble proteins, the supernatant was stored at -20°C, while the pellets were resuspended in 2ml buffer C (6M guanidine hydrochloride, 100mM phosphate buffer, 10mM Tris-HCl, pH 7.5) and treated in a homogenizer for 10 cycles. The product was centrifuged at 13000rpm for 40 minutes.

Supernatants were collected and mixed with 150µl Ni^{2+}-resin (Pharmacia) (previously washed with either buffer A or buffer B, as appropriate) and incubated at room temperature with gentle agitation for 30 minutes. The sample was centrifuged at 700*g* for 5 minutes at 4°C. The resin was washed twice with 10ml buffer A or B for 10 minutes, resuspended in 1ml buffer A or B and loaded on a disposable column. The resin was washed at either (i) 4°C with 2ml cold buffer A or (ii) room temperature with 2ml buffer B, until the flow-through reached OD_{280} of 0.02-0.06.

The resin was washed with either (i) 2ml cold 20mM imidazole buffer (300mM NaCl, 50mM phosphate buffer, 20mM imidazole, pH 8) or (ii) buffer D (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 6.3) until the flow-through reached the $O.D_{280}$ of 0.02-0.06. The His-fusion protein was eluted by addition of 700µl of either (i) cold elution buffer A (300mM NaCl, 50mM phosphate buffer, 250mM imidazole, pH 8) or (ii) elution buffer B (urea 8M, 10mM Tris-HCl, 100mM phosphate buffer, pH 4.5) and fractions collected until the $O.D_{280}$ was 0.1. 21µl of each fraction were loaded on a 12% SDS gel.

K) His-fusion proteins renaturation

10% glycerol was added to the denatured proteins. The proteins were then diluted to 20µg/ml using dialysis buffer I (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, 2M urea, pH 8.8) and dialysed against the same buffer at 4°C for 12-14 hours. The protein was further dialysed against dialysis buffer II (10% glycerol, 0.5M arginine, 50mM phosphate buffer, 5mM reduced glutathione, 0.5mM oxidised glutathione, pH 8.8) for 12-14 hours at 4°C. Protein concentration was evaluated using the formula:

$$\text{Protein (mg/ml)} = (1.55 \times OD_{280}) - (0.76 \times OD_{260})$$

L) His-fusion large-scale purification (ORFs 111-129)

500ml of bacterial cultures were induced and the fusion proteins were obtained soluble in buffer M1, M2 or M3 using the procedure described above. The crude extract of the bacteria was loaded onto a Ni-NTA superflow column (Quiagen) equilibrated with buffer M1, M2 or M3 depending on the solubilization buffer of the fusion proteins. Unbound material was eluted by washing the column with the same buffer. The specific protein was eluted with the corresponding buffer containing 500mM imidazole and dialysed against the corresponding buffer without imidazole. After each run the columns were sanitized by washing with at least two column volumes of 0.5 M sodium hydroxide and reequilibrated before the next use.

M) Mice immunisations

20μg of each purified protein were used to immunise mice intraperitoneally. In the case of ORFs 2, 4, 15, 22, 27, 28, 37, 76, 89 and 97, Balb-C mice were immunised with $Al(OH)_3$ as adjuvant on days 1, 21 and 42, and immune response was monitored in samples taken on day 56. For ORFs 44, 106 and 132, CD1 mice were immunised using the same protocol. For ORFs 25 and 40, CD1 mice were immunised using Freund's adjuvant, rather than $AL(OH)_3$, and the same immunisation protocol was used, except that the immune response was measured on day 42, rather than 56. Similarly, for ORFs 23, 32, 38 and 79, CD1 mice were immunised with Freund's adjuvant, but the immune response was measured on day 49.

N) ELISA assay (sera analysis)

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 7ml of Mueller-Hinton Broth (Difco) containing 0.25% Glucose. Bacterial growth was monitored every 30 minutes by following OD_{620}. The bacteria were let to grow until the OD reached the value of 0.3-0.4. The culture was centrifuged for 10 minutes at 10000rpm. The supernatant was discarded and bacteria were washed once with PBS, resuspended in PBS containing 0.025% formaldehyde, and incubated for 2 hours at room temperature and then overnight at 4°C with stirring. 100μl bacterial cells were added to each well of a 96 well Greiner plate and incubated overnight at 4°C. The wells were then washed three times with PBT washing buffer (0.1% Tween-20 in PBS). 200μl of saturation buffer (2.7% Polyvinylpyrrolidone 10 in water) was added to each well and the plates incubated for 2 hours at 37°C. Wells were washed

three times with PBT. 200μl of diluted sera (Dilution buffer: 1% BSA, 0.1% Tween-20, 0.1% NaN_3 in PBS) were added to each well and the plates incubated for 90 minutes at 37°C. Wells were washed three times with PBT. 100μl of HRP-conjugated rabbit anti-mouse (Dako) serum diluted 1:2000 in dilution buffer were added to each well and the plates were incubated for 90 minutes at 37°C. Wells were washed three times with PBT buffer. 100μl of substrate buffer for HRP (25ml of citrate buffer pH5, 10mg of O-phenildiamine and 10μl of H_2O) were added to each well and the plates were left at room temperature for 20 minutes. 100μl H_2SO_4 was added to each well and OD_{490} was followed. The ELISA was considered positive when OD_{490} was 2.5 times the respective pre-immune sera.

O) FACScan bacteria Binding Assay procedure.

The acapsulated MenB M7 strain was plated on chocolate agar plates and incubated overnight at 37°C. Bacterial colonies were collected from the agar plates using a sterile dracon swab and inoculated into 4 tubes containing 8ml each Mueller-Hinton Broth (Difco) containing 0.25% glucose. Bacterial growth was monitored every 30 minutes by following OD_{620}. The bacteria were let to grow until the OD reached the value of 0.35-0.5. The culture was centrifuged for 10 minutes at 4000rpm. The supernatant was discarded and the pellet was resuspended in blocking buffer (1% BSA, 0.4% NaN_3) and centrifuged for 5 minutes at 4000rpm. Cells were resuspended in blocking buffer to reach OD_{620} of 0.07. 100μl bacterial cells were added to each well of a Costar 96 well plate. 100μl of diluted (1:200) sera (in blocking buffer) were added to each well and plates incubated for 2 hours at 4°C. Cells were centrifuged for 5 minutes at 4000rpm, the supernatant aspirated and cells washed by addition of 200μl/well of blocking buffer in each well. 100μl of R-Phicoerytrin conjugated $F(ab)_2$ goat anti-mouse, diluted 1:100, was added to each well and plates incubated for 1 hour at 4°C. Cells were spun down by centrifugation at 4000rpm for 5 minutes and washed by addition of 200μl/well of blocking buffer. The supernatant was aspirated and cells resuspended in 200μl/well of PBS, 0.25% formaldehyde. Samples were transferred to FACScan tubes and read. The condition for FACScan setting were: FL1 on, FL2 and FL3 off; FSC-H threshold:92; FSC PMT Voltage: E 02; SSC PMT: 474; Amp. Gains 7.1; FL-2 PMT: 539; compensation values: 0.

P) OMV preparations

Bacteria were grown overnight on 5 GC plates, harvested with a loop and resuspended in 10 ml 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes and the bacteria disrupted by sonication for 10 minutes on ice (50% duty cycle, 50% output). Unbroken cells were removed by centrifugation at 5000*g* for 10 minutes and the total cell envelope fraction recovered by centrifugation at 50000*g* at 4°C for 75 minutes. To extract cytoplasmic membrane proteins from the crude outer membranes, the whole fraction was resuspended in 2% sarkosyl (Sigma) and incubated at room temperature for 20 minutes. The suspension was centrifuged at 10000*g* for 10 minutes to remove aggregates, and the supernatant further ultracentrifuged at 50000*g* for 75 minutes to pellet the outer membranes. The outer membranes were resuspended in 10mM Tris-HCl, pH8 and the protein concentration measured by the Bio-Rad Protein assay, using BSA as a standard.

Q) Whole Extracts preparation

Bacteria were grown overnight on a GC plate, harvested with a loop and resuspended in 1ml of 20mM Tris-HCl. Heat inactivation was performed at 56°C for 30 minutes.

R) Western blotting

Purified proteins (500ng/lane), outer membrane vesicles (5μg) and total cell extracts (25μg) derived from MenB strain 2996 were loaded on 15% SDS-PAGE and transferred to a nitrocellulose membrane. The transfer was performed for 2 hours at 150mA at 4°C, in transferring buffer (0.3 % Tris base, 1.44 % glycine, 20% methanol). The membrane was saturated by overnight incubation at 4°C in saturation buffer (10% skimmed milk, 0.1% Triton X100 in PBS). The membrane was washed twice with washing buffer (3% skimmed milk, 0.1% Triton X100 in PBS) and incubated for 2 hours at 37°C with mice sera diluted 1:200 in washing buffer. The membrane was washed twice and incubated for 90 minutes with a 1:2000 dilution of horseradish peroxidase labelled anti-mouse Ig. The membrane was washed twice with 0.1% Triton X100 in PBS and developed with the Opti-4CN Substrate Kit (Bio-Rad). The reaction was stopped by adding water.

S) Bactericidal assay

MC58 strain was grown overnight at 37°C on chocolate agar plates. 5-7 colonies were collected and used to inoculate 7ml Mueller-Hinton broth. The suspension was incubated at 37°C on a nutator and let to grow until OD_{620} was 0.5-0.8. The culture was aliquoted into sterile 1.5ml Eppendorf

tubes and centrifuged for 20 minutes at maximum speed in a microfuge. The pellet was washed once in Gey's buffer (Gibco) and resuspended in the same buffer to an OD_{620} of 0.5, diluted 1:20000 in Gey's buffer and stored at 25°C.

50µl of Gey's buffer/1% BSA was added to each well of a 96-well tissue culture plate. 25µl of diluted mice sera (1:100 in Gey's buffer/0.2% BSA) were added to each well and the plate incubated at 4°C. 25µl of the previously described bacterial suspension were added to each well. 25µl of either heat-inactivated (56°C waterbath for 30 minutes) or normal baby rabbit complement were added to each well. Immediately after the addition of the baby rabbit complement, 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 0). The 96-well plate was incubated for 1 hour at 37°C with rotation and then 22µl of each sample/well were plated on Mueller-Hinton agar plates (time 1). After overnight incubation the colonies corresponding to time 0 and time 1 hour were counted.

Table II (page 493) gives a summary of the cloning, expression and prurification results.

Example 1

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 1>:

```
  1  ATGAAACAGA CAGTCAA.AT GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
 51  GAACCGACCG GTGTGGNCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
101  A.GCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
151  TAT.TACAAA GGACGCGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
201  GTATCGGCAG CCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
251  GCTGGATGTA TGCCAACGGG CGCGC.GTGC GCCAAGATGA TACCGAAGCG
301  GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
351  CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGCGC CAAGACGATG
401  TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
451  GCCCAAAACA ATTTGGGCGT GATGTATGCC GAAAGANCGC GCGTGCGCCA
501  AGACCG...
```

This corresponds to the amino acid sequence <SEQ ID 2; ORF37>:

```
  1  MKQTVXMLAA ALIALGLNRP VWXDDVSDFR ENLXAAAQGN AAAQYNLGAM
 51  YXQRTRVRRD DAEAVRWYRQ PAEQGLAQAQ YNLGWMYANG RXVRQDDTEA
101  VRWYRQAAAQ GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAQ
151  AQNNLGVMYA ERXRVRQD...
```

Further work revealed the complete nucleotide sequence <SEQ ID 3>:

```
  1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
 51  GAACCGAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
101  AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AATACAATTT GGGCGCAATG
151  TATTACAAAG GACGCGGCGT GCGCCGGGAT GATGCTGAAG CGGTCAGATG
201  GTATCGGCAG GCGGCGGAAC AGGGGTTAGC CCAAGCCCAA TACAATTTGG
251  GCTGGATGTA TGCCAACGGG CGCGGCGTGC GCCAAGATGA TACCGAAGCG
301  GTCAGATGGT ATCGGCAGGC GGCAGCGCAG GGGGTTGTCC AAGCCCAATA
351  CAATTTGGGC GTGATATATG CCGAAGGACG TGGAGTGCGC CAAGACGATG
401  TCGAAGCGGT CAGATGGTTT CGGCAGGCGG CAGCGCAGGG GGTAGCCCAA
451  GCCCAAAACA ATTTGGGCGT GATGTATGCC GAAAGACGCG GCGTGCGCCA
501  AGACCGCGCC CTTGCACAAG AATGGTTTGG CAAGGCTTGT CAAAACGGAG
551  ACCAAGACGG CTGCGACAAT GACCAACGCC TGAAGGCGGG TTATTGA
```

This corresponds to the amino acid sequence <SEQ ID 4; ORF37-1>:

```
  1  MKQTVKWLAA ALIALGLNRA VWADDVSDFR ENLQAAAQGN AAAQYNLGAM
 51  YYKGRGVRRD DAEAVRWYRQ AAEQGLAQAQ YNLGWMYANG RGVRQDDTEA
101  VRWYRQAAAQ GVVQAQYNLG VIYAEGRGVR QDDVEAVRWF RQAAAQGVAQ
151  AQNNLGVMYA ERRGVRQDRA LAQEWFGKAC QNGDQDGCDN DQRLKAGY*
```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 5>:

```
  1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
 51  GAACCAAGCG GTGTGGGCGG ATGACGTATC GGATTTTCGG GAAAACTTGC
101  AGGCGGCAGC ACAGGGAAAT GCAGCAGCCC AAAACAATTT GGGCGTGATG
151  TATGCCGAAA GACGCGGCGT GCGCCAAGAC CGCGCCCTTG CACAAGAATG
201  GCTTGGCAAG GCTTGTCAAA ACGGATACCA AGACAGCTGC GACAATGACC
251  AACGCCTGAA AGCGGGTTAT TGA
```

This encodes a protein having amino acid sequence <SEQ ID 6; ORF37a>:

```
  1  MKQTVKWLAA ALIALGLNQA VWADDVSDFR ENLQAAAQGN AAAQNNLGVM
 51  YAERRGVRQD RALAQEWLGK ACQNGYQDSC DNDQRLKAGY *
```

The originally-identified partial strain B sequence (ORF37) shows 68.0% identity over a 75aa overlap with ORF37a:

```
                      10        20        30        40        50        60
orf37.pep    MKQTVXMLAAALIALGLNRPVWXDDVSDFRENLXAAAQGNAAAQYNLGAMYXQRTRVRRD
             |||||  |||||||||||: ||  |||||||||| |||||||||| |||:|| :|  ||:|
orf37a       MKQTVKWLAAALIALGLNQAVWADDVSDFRENLQAAAQGNAAAQNNLGVMYAERRGVRQD
                      10        20        30        40        50        60

                      70        80        90       100       110       120
orf37.pep    DAEAVRWYRQPAEQGLAQAQYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG
              | | :|  :  ::|
orf37a       RALAQEWLGKACQNGYQDSCDNDQRLKAGYX
                      70        80        90
```

Further work identified the corresponding gene in *N.gonorrhoeae* <SEQ ID 7 >:

```
  1  ATGAAACAGA CAGTCAAATG GCTTGCCGCC GCCCTGATTG CCTTGGGCTT
 51  GAACCAAGCG GTGTGGGCGG GTGACGTATC GGATTTTCGG GAAAACTTGC
101  AGgcggcaGA ACaggGAAAT GCAGCAGCCC AATTCAATTT GGGCGTGATG
151  TATGAAAATG GACAAGGAGT TCGTCAAGAT TATGTACAGG CAGTGCAGTG
201  GTATCGCAAG GCTTCAGAAC AAGGGGATGC CCAAGCCCAA TACAATTTGG
251  GCTTGATGTA TTACGATGGA CGCGGCGTGC GCCAAGACCT TGCGCTCGCT
301  CAACAATGGC TTGGCAAGGC TTGTCAAAAC GGAGACCAAA ACAGCTGCGA
351  CAATGACCAA CGCCTGAAGG CGGGTTATTA A
```

This encodes a protein having amino acid sequence <SEQ ID 8; ORF37ng>:

```
  1  MKQTVKWLAA ALIALGLNQA VWAGDVSDFR ENLQAAEQGN AAAQFNLGVM
 51  YENGQGVRQD YVQAVQWYRK ASEQGDAQAQ YNLGLMYYDG RGVRQDLALA
101  QQWLGKACQN GDQNSCDNDQ RLKAGY*
```

The originally-identified partial strain B sequence (ORF37) shows 64.9% identity over a 111aa overlap with ORF37ng:

```
orf37.pep    MKQTVXMLAAALIALGLNRPVWXDDVSDFRENLXAAAQGNAAAQYNLGAMYXQRTRVRRD   60
             |||||  |||||||||||: ||  ||||||||| || |||||||:|||:||   :  ||:|
orf37ng      MKQTVKWLAAALIALGLNQAVWAGDVSDFRENLQAAEQGNAAAQFNLGVMYENGQGVRQD   60

orf37.pep    DAEAVRWYRQPAEQGLAQAQYNLGWMYANGRXVRQDDTEAVRWYRQAAAQGVVQAQYNLG  120
              ::||:|||: :||| |||||||| || :|| |||| : | :|  :|  :|
orf37ng      YVQAVQWYRKASEQGDAQAQYNLGLMYYDGRGVRQDLALAQQWLGKACQNGDQNSCDNDQ  120

orf37.pep    VIYAEGRGVRQDDVEAVRWFRQAAAQGVAQAQNNLGVMYAERXRVRQD  168

orf37ng      RLKAGY                                            126
```

The complete strain B sequence (ORF37-1) and ORF37ng show 51.5% identity in 198 aa overlap:

```
                          10        20        30        40        50        60
orf37-1.pep       MKQTVKWLAAALIALGLNRAVWADDVSDFRENLQAAAQGNAAAQYNLGAMYYKGRGVRRD
                  ||||||||||||||||||:|||| |||||||||||| |||||||:|||:|| :|:|||:|
orf37ng           MKQTVKWLAAALIALGLNQAVWAGDVSDFRENLQAAEQGNAAAQFNLGVMYENGQGVRQD
                          10        20        30        40        50        60

                          70        80        90       100       110       120
orf37-1.pep       DAEAVRWYRQAAEQGLAQAQYNLGWMYANGRGVRQDDTEAVRWYRQAAAQGVVQAQYNLG
                   ::||:|||:|:||| |||||||| || :|||||||
orf37ng           YVQAVQWYRKASEQGDAQAQYNLGLMYYDGRGVRQD------------------------
                          70        80        90

                         130       140       150       160       170       180
orf37-1.pep       VIYAEGRGVRQDDVEAVRWFRQAAAQGVAQAQNNLGVMYAERRGVRQDRALAQEWFGKAC
                                                                   ||||:|:||||
orf37ng           ------------------------------------------------LALAQQWLGKAC
                                                                   100

                         190       199
orf37-1.pep       QNGDQDGCDNDQRLKAGYX
                  |||||::||||||||||||
orf37ng           QNGDQNSCDNDQRLKAGYX
                 110       120
```

Computer analysis of these amino acid sequences indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF37-1 (11kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 1A shows the results of affinity purification of the GST-fusion protein, and Figure 1B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), FACS analysis (Figure 1C), and a bactericidal assay (Figure 1D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 1E shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF37-1.

Example 2

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 9>:

```
  TTCGGCGA CATCGGCGGT TTGAAGGTCA ATGCCCCCGT CAAATCCGCA
GGCGTATTGG TCGGGCGCGT CGGCGCTATC GGACTTGACC CGAAATCCTA
TCAGGCGAGG GTGCGCCTCG ATTTGGACGG CAAGTATCAG TTCAGCAGCG
ACGTTTCCGC GCAAATCCTG ACTTCsGGAC TTTTGGGCGA GCAGTACATC
GGGCTGCAGC AGGGCGGCGA CACGGAAAAC CTTGCTGCCG GCGACACCAT
CTCCGTAACC AGTTCTGCAA TGGTTCTGGA AAACCTTATC GGCAAATTCA
TGACGAGTTT TGCCGAGAAA AATGCCGACG GCGGCAATGC GGAAAAAGCC
GCCGAATAA
```

This corresponds to the amino acid sequence <SEQ ID 10>:

```
 1  FGDIGGLKVN APVKSAGVLV GRVGAIGLDP KSYQARVRLD LDGKYQFSSD
51  VSAQILTSGL LGEQYIGLQQ GGDTENLAAG DTISVTSSAM VLENLIGKFM
```

```
101  TSFAEKNADG GNAEKAAE*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a hypothetical *H.influenzae* protein (ybrd.haein; accession number p45029)

SEQ ID 9 and ybrd.haein show 48.4% aa identity in 122 aa overlap:

```
                20        30        40        50        60        70
yrbd.h  LGIGALVFLGLRVANVQGFAETKSYTVTATFDNIGGLKVRAPLKIGGVVIGRVSAITLDE
                                            |::|||||||:||:| :||::|||:||:||
N.m                                         FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
                                                    10        20        30

                80        90       100       110       120       130
yrbd.h  KSYLPKVSIAINQEYNEIPENSSLSIKTSGLLGEQYIALTMGFDDGDTAMLKNGSQIQDT
        ||| ::|:::::: :| :::::  |    |  |||||||||||:|  |      |||: | :|: |  |
N.m     KSYQARVRLDLDGKY-QFSSDVSAQILTSGLLGEQYIGLQQG---GDTENLAAGDTISVT
                   40        50        60        70            80

               140       150       160
yrbd.h  TSAMVLEDLIGQFL--YGSKKSDGNEKSESTEQ
        :||||||:|||:|:  :::|::||:: ::::|:
N.m     SSAMVLENLIGKFMTSFAEKNADGGNAEKAAEX
             90       100       110       120
```

Homology with a predicted ORF from *N.gonorrhoeae*

SEQ ID 9 shows 99.2% identity over a 118aa overlap with a predicted ORF from *N. gonorrhoeae:*

```
                      20        30        40        50        60        70
yrbd            GAAAVAFLAFRVAGGAAFGGSDKTYAVYADFGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
                                                ||||||||||||||||||||||||||||||
N.m                                             FGDIGGLKVNAPVKSAGVLVGRVGAIGLDP
                                                        10        20        30

                      80        90       100       110       120       130
yrbd            KSYQARVRLDLDGKYQFSSDVSAQILTSGLLGEQYIGLQQGGDTENLAAGDTISVTSSAM
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
N.m             KSYQARVRLDLDGKYQFSSDVSAQILTSGLLGEQYIGLQQGGDTENLAAGDTISVTSSAM
                          40        50        60        70        80        90

                     140       150       160
yrbd            VLENLIGKFMTSFAEKNAEGGNAEKAAEX
                |||||||||||||||||:||||||||||
N.m             VLENLIGKFMTSFAEKNADGGNAEKAAEX
                          100       110       120
```

The complete yrbd *H.influenzae* sequence has a leader sequence and it is expected that the full-length homologous *N.meningitidis* protein will also have one. This suggests that it is either a membrane protein, a secreted protein, or a surface protein and that the protein, or one of its epitopes, could be a useful antigen for vaccines or diagnostics.

Example 3

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 11>:

```
  1  ..ATTTTGATAT ACCTCATCCG CAAGAATCTA GGTTCGCCCG TCTTCTTCTT
 51    TCAGGAACGC CCCGGAAAGG ACGGAAAACC TTTTAAAATG GTCAAATTCC
101    GTTCCATGCG CGACGGCTTG TATTCAGACG GCATTCCGCT GCCCGACGGA
151    GAACGCCTGA CACCGTTCGG CAAAAAACTG CGTGCCGcCA GTwTGGACGA
201    ACTGCCTGAA TTATGGAATA TCTTAAAAGG CGAGATGAGC CTGGTCGGCC
251    CCCGCCCGCT GCTGATGCAA TATCTGCCGC TGTACGACAA CTTCCAAAAC
301    CGCCGCCACG AAATGAAACC CGGCATTACC GGCTGGGCGC AGGTCAACGG
```

```
351    GCGCAACGCg CTTTCGTGGG ACGAAAAATT CGCCTGCGAT GTTTGGTATA
401    TCGACCACTT CAGCCTGTGC CTCGACATCA AAATCCTACT GCTGACGGTT
451    AAAAAAGTAT TAATCAAGGA AGGGATTTCC GCACAGGGCG AACA.aCCAT
501    GCCCCCTTTC ACAGGAAAAC GCAAACTCGC CGTCGTCGGT GCGGGCGGAC
551    ACGGAAAAGT CGTTGCCGAC CTTGCCGCCG CACTCGGCCG GTACAGGGAA
601    ATCGTTTTTC TGGACGACCG CGCACAAGGC AGCGTCAACG GCTTTTCCGT
651    CATCGGCACG ACGCTGCTGC TTGAAAACAG TTTATCGCCC GAACAATACG
701    ACGTCGCCGT CGCCGTCGGC AACAACCGCA TCCGCCGCCA AATCGCCGAA
751    AAAGCCGCCG CGCTCGGCTT CGCCCTGCCC GTACTGGTTC ATCCGGACGC
801    GACCGTCTCG CCTTCTGCAA CAGTCGGACA AGGCAGCGTC GTTATGGCGA
851    AAGCGGTCG..
```

This corresponds to the amino acid sequence <SEQ ID 12; ORF3>:

```
  1  ..ILIYLIRKNL GSPVFFFQER PGKDGKPFKM VKFRSMRDGL YSDGIPLPDG
 51    ERLTPFGKKL RAASXDELPE LWNILKGEMS LVGPRPLLMQ YLPLYDNFQN
101    RRHEMKPGIT GWAQVNGRNA LSWDEKFACD VWYIDHFSLC LDIKILLLTV
151    KKVLIKEGIS AQGEXTMPPF TGKRKLAVVG AGGHGKVVAD LAAALGRYRE
201    IVFLDDRAQG SVNGFSVIGT TLLLENSLSP EQYDVAVAVG NNRIRRQIAE
251    KAAALGFALP VLVHPDATVS PSATVGQGSV VMAKAV..
```

Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 13>:

```
   1  ATGAGTAAAT TCTTCAAACG CCTGTTTGAC ATTGTTGCCT CCGCCTCGGG
  51  ACTGATTTTC CTCTCGCCAG TATTTTTGAT TTTGATATAC CTCATCCGCA
 101  AGAATCTAGG TTCGCCCGTC TTCTTCTTTC AGGAACGCCC CGGAAAGGAC
 151  GGAAAACCTT TTAAAATGGT CAAATTCCGT TCCATGCGCG ACGCGCTTGA
 201  TTCAGACGGC ATTCCGCTGC CCGACGGAGA ACGCCTGACA CCGTTCGGCA
 251  AAAAACTGCG TGCCGCCAGT TTGGACGAAC TGCCTGAATT ATGGAATATC
 301  TTAAAAGGCG AGATGAGCCT GGTCGGCCCC CGCCCGCTGC TGATGCAATA
 351  TCTGCCGCTG TACGACAACT TCCAAAACCG CCGCCACGAA ATGAAACCCG
 401  GCATTACCGG CTGGGCGCAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC
 451  GAAAAATTCG CCTGCGATGT TTGGTATATC GACCACTTCA GCCTGTGCCT
 501  CGACATCAAA ATCCTACTGC TGACGGTTAA AAAAGTATTA ATCAAGGAAG
 551  GGATTTCCGC ACAGGGCGAA GCCACCATGC CCCCTTTCAC AGGAAAACGC
 601  AAACTCGCCG TCGTCGGTGC GGGCGGACAC GGAAAAGTCG TTGCCGACCT
 651  TGCCGCCGCA CTCGGCCGGT ACAGGGAAAT CGTTTTTCTG GACGACCGCG
 701  CACAAGGCAG CGTCAACGGC TTTTCCGTCA TCGGCACGAC GCTGCTGCTT
 751  GAAAACAGTT TATCGCCCGA ACAATACGAC GTCGCCGTCG CCGTCGGCAA
 801  CAACCGCATC CGCCGCCAAA TCGCCGAAAA AGCCGCCGCG CTCGGCTTCG
 851  CCCTGCCCGT TCTGGTTCAT CCGGACGCGA CCGTCTCGCC TTCTGCAACA
 901  GTCGGACAAG GCAGCGTCGT TATGGCGAAA GCCGTCGTAC AGGCAGGCAG
 951  CGTATTGAAA GACGGCGTGA TTGTGAACAC TGCCGCCACC GTCGATCACG
1001  ACTGCCTGCT TAACGCTTTC GTCCACATCA GCCCAGGCGC GCACCTGTCG
1051  GGCAACACGC ATATCGGCGA AGAAAGCTGG ATAGGCACGG GCGCGTGCAG
1101  CCGCCAGCAG ATCCGTATCG GCAGCCGCGC AACCATTGGA GCGGGCGCAG
1151  TCGTCGTACG CGACGTTTCA GACGGCATGA CCGTCGCGGG CAATCCGGCA
1201  AAGCCGCTGC CGCGCAAAAA CCCCGAGACC TCGACAGCAT AA
```

This corresponds to the amino acid sequence <SEQ ID 14; ORF3-1>:

```
  1  MSKFFKRLFD IVASASGLIF LSPVFLILIY LIRKNLGSPV FFFQERPGKD
 51  GKPFKMVKFR SMRDALDSDG IPLPDGERLT PFGKKLRAAS LDELPELWNI
101  LKGEMSLVGP RPLLMQYLPL YDNFQNRRHE MKPGITGWAQ VNGRNALSWD
151  EKFACDVWYI DHFSLCLDIK ILLLTVKKVL IKEGISAQGE ATMPPFTGKR
201  KLAVVGAGGH GKVVADLAAA LGRYREIVFL DDRAQGSVNG FSVIGTTLLL
251  ENSLSPEQYD VAVAVGNNRI RRQIAEKAAA LGFALPVLVH PDATVSPSAT
301  VGQGSVVMAK AVVQAGSVLK DGVIVNTAAT VDHDCLLNAF VHISPGAHLS
351  GNTHIGEESW IGTGACSRQQ IRIGSRATIG AGAVVVRDVS DGMTVAGNPA
401  KPLPRKNPET STA*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF3 shows 93.0% identity over a 286aa overlap with an ORF (ORF3a) from strain A of *N. meningitidis*:

```
                                                      10        20        30
```

```
orf3.pep                             ILIYLIRKNLGSPVFFFQERPGKDGKPFKMVKFR
                                     ||||||||||||||||||||||||||||||||||
orf3a      MSKFFKRLFDIVASASGLIFLSPVFLILIYLIRKNLGSPVFFFQERPGKDGKPFKMVKFR
                    10        20        30        40        50        60

                   40        50        60        70        80        90
orf3.pep   SMRDGLYSDGIPLPDGERLTPFGKKLRAASXDELPELWNILKGEMSLVGPRPLLMQYLPL
           ||:|:| |||| |||||||||||||||||| ||||||||:|||:||||||||||||||||
orf3a      SMHDALDSDGILLPDGERLTPFGKKLRAASLDELPELWNVLKGDMSLVGPRPLLMQYLPL
                    70        80        90       100       110       120

                  100       110       120       130       140       150
orf3.pep   YDNFQNRRHEMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFSLCLDIKILLLTVKKVL
           |||||||||||||||||||||||||||||||:||||:|||||||||||||||||||||||
orf3a      YDNFQNRRHEMKPGITGWAQVNGRNALSWDERFACDIWYIDHFSLCLDIKILLLTVKKVL
                   130       140       150       160       170       180

                  160       170       180       190       200       210
orf3.pep   IKEGISAQGEXTMPPFTGKRKLAVVGAGGHGKVVADLAAALGRYREIVFLDDRAQGSVNG
           |||||||||| ||||||||||||||||||||||||:|||||| | ||||||||:||||||
orf3a      IKEGISAQGEATMPPFTGKRKLAVVGAGGHGKVVAELAAALGTYGEIVFLDDRVQGSVNG
                   190       200       210       220       230       240

                  220       230       240       250       260       270
orf3.pep   FSVIGTTLLLENSLSPEQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT
           | ||||||||||||||||:|:|||||||||||||||||||||||||||:|||:|||||||
orf3a      FPVIGTTLLLENSLSPEQFDIAVAVGNNRIRRQIAEKAAALGFALPVLIHPDSTVSPSAT
                   250       260       270       280       290       300

                  280
orf3.pep   VGQGSVVMAKAV
           ||||:|||||||
orf3a      VGQGGVVMAKAVVQADSVLKDGVIVNTAATVDHDCLLDAFVHISPGAHLSGNTRIGEESW
                   310       320       330       340       350       360
```

The complete length ORF3a nucleotide sequence <SEQ ID 15> is:

```
   1  ATGAGTAAAT TCTTCAAACG CCTGTTTGAC ATTGTTGCCT CCGCCTCGGG
  51  ACTGATTTTC CTCTCGCCAG TATTTTTGAT TTTGATATAC CTCATCCGCA
 101  AGAATCTGGG TTCGCCCGTC TTCTTCTTTC AGGAACGCCC CGGAAAGGAC
 151  GGAAAACCTT TTAAAATGGT CAAATTCCGT TCCATGCACG ACGCGCTTGA
 201  TTCAGACGGC ATTCTGCTGC CCGACGGAGA ACGCCTGACA CCGTTCGGCA
 251  AAAAACTGCG TGCCGCCAGT TTGGACGAAC TGCCCGAACT GTGGAACGTC
 301  CTCAAAGGCG ACATGAGCCT GGTCGGCCCC CGCCCGCTGC TGATGCAATA
 351  TCTGCCGCTG TACGACAACT TCCAAAACCG CCGCCACGAA ATGAAACCGG
 401  GCATTACCGG CTGGGCGCAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC
 451  GAACGCTTCG CATGCGACAT CTGGTATATC GACCACTTCA GCCTGTGCCT
 501  CGACATCAAA ATCCTACTGC TGACGGTTAA AAAAGTATTA ATCAAAGAAG
 551  GGATTTCCGC ACAGGGCGAA GCCACCATGC CCCCTTTCAC AGGAAAACGC
 601  AAACTTGCCG TCGTCGGTGC GGGCGGACAC GGCAAAGTCG TTGCCGAGCT
 651  TGCCGCCGCA CTCGGCACAT ACGGCGAAAT CGTTTTTCTG GACGACCGCG
 701  TCCAAGGCAG CGTCAACGGC TTCCCCGTCA TCGGCACGAC GCTGCTGCTT
 751  GAAAACAGTT TATCGCCCGA ACAATTCGAC ATCGCCGTCG CCGTCGGCAA
 801  CAACCGCATC CGCCGCCAAA TCGCCGAAAA AGCCGCCGCG CTCGGCTTCG
 851  CCCTGCCCGT CCTGATTCAT CCGGACTCGA CCGTCTCGCC TTCTGCAACA
 901  GTCGGACAAG GCGGCGTCGT TATGGCGAAA GCCGTCGTAC AGGCTGACAG
 951  CGTATTGAAA GACGGCGTAA TTGTGAACAC TGCCGCCACC GTCGATCACG
1001  ATTGCCTGCT TGATGCTTTC GTCCACATCA GCCCGGGCGC GCACCTGTCG
1051  GGCAACACGC GTATCGGCGA AGAAAGCTGG ATAGGCACAG GCGCGTGCAG
1101  CCGCCAGCAG ATCCGTATCG GCAGCCGCGC AACCATTGGA GCGGGCGCAG
1151  TCGTCGTGCG CGACGTTTCA GACGGCATGA CCGTCGCGGG CAACCCGGCA
1201  AAACCATTGG CAGGCAAAAA TACCGAGACC CTGCGGTCGT AA
```

This is predicted to encode a protein having amino acid sequence <SEQ ID 16>:

```
   1  MSKFFKRLFD IVASASGLIF LSPVFLILIY LIRKNLGSPV FFFQERPGKD
  51  GKPFKMVKFR SMHDALDSDG ILLPDGERLT PFGKKLRAAS LDELPELWNV
 101  LKGDMSLVGP RPLLMQYLPL YDNFQNRRHE MKPGITGWAQ VNGRNALSWD
 151  ERFACDIWYI DHFSLCLDIK ILLLTVKKVL IKEGISAQGE ATMPPFTGKR
 201  KLAVVGAGGH GKVVAELAAA LGTYGEIVFL DDRVQGSVNG FPVIGTTLLL
 251  ENSLSPEQFD IAVAVGNNRI RRQIAEKAAA LGFALPVLIH PDSTVSPSAT
```

```
301  VGQGGVVMAK AVVQADSVLK DGVIVNTAAT VDHDCLLDAF VHISPGAHLS
351  GNTRIGEESW IGTGACSRQQ IRIGSRATIG AGAVVVRDVS DGMTVAGNPA
401  KPLAGKNTET LRS*
```

Two transmembrane domains are underlined.

ORF3-1 shows 94.6% identity in 410 aa overlap with ORF3a:

```
                      10        20        30        40        50        60
orf3a.pep     MSKFFKRLFDIVASASGLIFLSPVFLILIYLIRKNLGSPVFFFQERPGKDGKPFKMVKFR
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf3-1        MSKFFKRLFDIVASASGLIFLSPVFLILIYLIRKNLGSPVFFFQERPGKDGKPFKMVKFR
                      10        20        30        40        50        60

                      70        80        90       100       110       120
orf3a.pep     SMHDALDSDGILLPDGERLTPFGKKLRAASLDELPELWNVLKGDMSLVGPRPLLMQYLPL
              ||:|||||||| |||||||||||||||||||||||||||:|||:||||||||||||||||
orf3-1        SMRDALDSDGIPLPDGERLTPFGKKLRAASLDELPELWNILKGEMSLVGPRPLLMQYLPL
                      70        80        90       100       110       120

                     130       140       150       160       170       180
orf3a.pep     YDNFQNRRHEMKPGITGWAQVNGRNALSWDERFACDIWYIDHFSLCLDIKILLLTVKKVL
              |||||||||||||||||||||||||||||||:||||:|||||||||||||||||||||||
orf3-1        YDNFQNRRHEMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFSLCLDIKILLLTVKKVL
                     130       140       150       160       170       180

                     190       200       210       220       230       240
orf3a.pep     IKEGISAQGEATMPPFTGKRKLAVVGAGGHGKVVAELAAALGTYGEIVFLDDRVQGSVNG
              |||||||||||||||||||||||||||||||||||:|||||| | ||||||||:||||||
orf3-1        IKEGISAQGEATMPPFTGKRKLAVVGAGGHGKVVADLAAALGRYREIVFLDDRAQGSVNG
                     190       200       210       220       230       240

                     250       260       270       280       290       300
orf3a.pep     FPVIGTTLLLENSLSPEQFDIAVAVGNNRIRRQIAEKAAALGFALPVLIHPDSTVSPSAT
              | ||||||||||||||||:|:|||||||||||||||||||||||||||:|||:|||||||
orf3-1        FSVIGTTLLLENSLSPEQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT
                     250       260       270       280       290       300

                     310       320       330       340       350       360
orf3a.pep     VGQGGVVMAKAVVQADSVLKDGVIVNTAATVDHDCLLDAFVHISPGAHLSGNTRIGEESW
              ||||:|||||||||| |||||||||||||||||||||:|||||||||||||||:||||||
orf3-1        VGQGSVVMAKAVVQAGSVLKDGVIVNTAATVDHDCLLNAFVHISPGAHLSGNTHIGEESW
                     310       320       330       340       350       360

                     370       380       390       400       410
orf3a.pep     IGTGACSRQQIRIGSRATIGAGAVVVRDVSDGMTVAGNPAKPLAGKNTETLRSX
              |||||||||||||||||||||||||||||||||||||||||||  || ||
orf3-1        IGTGACSRQQIRIGSRATIGAGAVVVRDVSDGMTVAGNPAKPLPRKNPETSTAX
                     370       380       390       400       410
```

<u>Homology with hypothetical protein encoded by *yvfc* gene (accession Z71928) of *B. subtilis*</u>

ORF3 and YVFC proteins show 55% aa identity in 170 aa overlap (BLASTp):

```
ORF3    3   IYLIRKNLGSPVFFFQERPGKDGKPFKMVKFRSMRDGLYSDGIPLPDGERLTPFGKKLRA 62
            I ++R  +GSPVFF Q RPG  GKPF + KFR+M D   S G  LPD  RLT  G+ +R
yvfc   27   IAVVRLKIGSPVFFKQVRPGLHGKPFTLYKFRTMTDERDSKGNLLPDEVRLTKTGRLIRK 86

ORF3   63   ASXDELPELWNILKGEMSLVGPRPLLMQYLPLYDNFQNRRHEMKPGITGWAQVNGRNALS 122
             S DELP+L N+LKG++SLVGPRPLLM YLPLY    Q RRHE+KPGITGWAQ+NGRNA+S
yvfc   87   LSIDELPQLLNVLKGDLSLVGPRPLLMDYLPLYTEKQARRHEVKPGITGWAQINGRNAIS 146

ORF3   123 WDEKFACDVWYIDHFSLCLDXXXXXXXXXXXXXXXEGISAQGEXTMPPFTG 172
           W++KF  DVWY+D++S  LD                EGI       T   FTG
yvfc   147 WEKKFELDVWYVDNWSFFLDLKILCLTVRKVLVSEGIQQTNHVTAERFTG 196
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF3 shows 86.3% identity over a 286aa overlap with a predicted ORF (ORF3.ng) from *N. gonorrhoeae*:

```
orf3                                ILIYLIRKNLGSPVFFFQERPGKDGKPFKMVKFR   34
                                    :|||||||| ||||||::||||||||||||||||
orf3ng    MSKAVKRLFDIIASASGLIVLSPVFLVLIYLIRKNKGSPVFFIRERPGKDGKPFKMVKFR   60

orf3      SMRDGLYSDGIPLPDGERLTPFGKKLRAASXDELPELWNILKGEMSLVGPRPLLMQYLPL   94
          ||||:| ||||||||:|||| |||||||:| ||||||||:||||||||||||||||||||
orf3ng    SMRDALDSDGIPLPDSERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRPLLMQYLPL  120

orf3      YDNFQNRRHEMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFSLCLDIKILLLTVKKVL  154
          |::||||||||||||||||||||||||||||||:||||| |:||: ||:|||:|||||||
orf3ng    YNKFQNRRHEMKPGITGWAQVNGRNALSWDEKFSCDVWYTDNFSFWLDMKILFLTVKKVL  180

orf3      IKEGISAQGEXTMPPFTGKRKLAVVGAGGHGKVVADLAAALGRYREIVFLDDRAQGSVNG  214
          |||||||||| |||||:|:|||||:||||||||||:|||||| | ||||||||:||||||
orf3ng    IKEGISAQGEATMPPFAGNRKLAVIGAGGHGKVVAELAAALGTYGEIVFLDDRTQGSVNG  240

orf3      FSVIGTTLLLENSLSPEQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT  274
          | ||||||||||||||||:|::||||||||||||:|:|||||| ||||:||||||||||
orf3ng    FPVIGTTLLLENSLSPEQFDITVAVGNNRIRRQITENAAALGFKLPVLIHPDATVSPSAI  300

orf3      VGQGSVVMAKAV                                                  286
          :|||||||||||
orf3ng    IGQGSVVMAKAVVQAGSVLKDGVIVNTAATVDHDCLLDAFVHISPGAHLSGNTRIGEESR  360
```

The complete length ORF3ng nucleotide sequence <SEQ ID 17> is:

```
   1  ATGAGTAAAG CCGTCAAACG CCTGTTCGAC ATCATCGCAT CCGCATCGGG
  51  GCTGATTGTC CTGTCGCCCG TGTTTTTGGT TTTAATATAC CTCATCCGCA
 101  AAAACTTAGG TTCGCCCGTC TTCTTCattC GGGAACGCCc cgGAAAGGAc
 151  ggaaaacCTT TTAAAATGGT CAAATTCCGT TCCAtgcgcg acgcgcttGA
 201  TTCAGACGGC ATTCCGCTGC CCGATAGCGA ACGCCTGACC GATTTCGGCA
 251  AAAAATTACG CGCCACCAGT TTGGACGAAC TTCCTGAATT ATGGAATGTC
 301  CTCAAAGGCG AGATGAGCCT GGTCGGCCCC CGCCCGCTTT TGATGCAGTA
 351  TCTGCCGCTT TACAACAAAT TTCAAAACCG CCGCCACGAA ATGAAACCGG
 401  GCATTACCGG CTGGGCGCAG GTCAACGGGC GCAACGCGCT TTCGTGGGAC
 451  GAAAAGTTCT CCTGCGATGT TTGGTACACC GACAATTTCA GCTTTTGGCT
 501  GGATATGAAA ATCCTGTTTC TGACAGTCAA AAAAGTCTTG ATTAAAGAAG
 551  GCATTTCGGC GCAAGGGGAA GCCACCATGC CCCCTTTCGC GGGGAATCGC
 601  AAACTCGCCG TTATCGGCGC GGGCGGACAC GGCAAAGTCG TTGCCGAGCT
 651  TGCCGCCGCA CTCGGCACAT ACGGCGAAAT CGTTTTTCTG GACGACCGCA
 701  CCCAAGGCAG CGTCAACGGC TTCCCCGTCA TCGGCACGAC GCTGCTGCTT
 751  GAAAACAGTT TATCGCCCGA ACAATTCGAC ATCACCGTCG CCGTCGGCAA
 801  CAACCGCATC CGCCGCCAAA TCACCGAAAA CGCCGCCGCG CTCGGCTTCA
 851  AACTGCCCGT TCTGATTCAT CCCGACGCGA CCGTCTCGCC TTCTGCAATA
 901  ATCGGACAAG GCAGCGTCGT AATGGCGAAA GCCGTCGTAC AGGCCGGCAG
 951  CGTATTGAAA GACGGCGTGA TTGTGAACAC TGCCGCCACC GTCGATCACG
1001  ACTGCCTGCT TGACGCTTTC GtccaCATCA GCCCGGGCGC GCACCTGTCG
1051  GGCAACACGC GTATCGGCGA AGAAAGCCGG ATAGGCACGG GCGCGTGCAG
1101  CCGCCAGCAG ACAACCGTCG GCAGCGGGGT TACCgccgGT GCAGGGgcGG
1151  TTATCGTATG CGACATCCCG GACGGCATGA CCGTCGCGGG CAACCCGGCA
1201  AAGCCCCTTA CGGGCAAAAA CCCCAAGACC GGGACGGCAT AA
```

This encodes a protein having amino acid sequence <SEQ ID 18>:

```
  1  MSKAVKRLFD IIASASGLIV LSPVFLVLIY LIRKNLGSPV FFIRERPGKD
 51  GKPFKMVKFR SMRDALDSDG IPLPDSERLT DFGKKLRATS LDELPELWNV
101  LKGEMSLVGP RPLLMQYLPL YNKFQNRRHE MKPGITGWAQ VNGRNALSWD
151  EKFSCDVWYT DNFSFWLDMK ILFLTVKKVL IKEGISAQGE ATMPPFAGNR
201  KLAVIGAGGH GKVVAELAAA LGTYGEIVFL DDRTQGSVNG FPVIGTTLLL
251  ENSLSPEQFD ITVAVGNNRI RRQITENAAA LGFKLPVLIH PDATVSPSAI
301  IGQGSVVMAK AVVQAGSVLK DGVIVNTAAT VDHDCLLDAF VHISPGAHLS
351  GNTRIGEESR IGTGACSRQQ TTVGSGVTAG AGAVIVCDIP DGMTVAGNPA
401  KPLTGKNPKT GTA*
```

This protein shows 86.9% identity in 413 aa overlap with ORF3-1:

```
                       10        20        30        40        50        60
orf3-1.pep     MSKFFKRLFDIVASASGLIFLSPVFLILIYLIRKNLGSPVFFFQERPGKDGKPFKMVKFR
               |||  ||||||:||||||| ||||||:|||||||||||||||::||||||||||||||||
orf3ng         MSKAVKRLFDIIASASGLIVLSPVFLVLIYLIRKNLGSPVFFIRERPGKDGKPFKMVKFR
                       10        20        30        40        50        60

                       70        80        90       100       110       120
orf3-1.pep     SMRDALDSDGIPLPDGERLTPFGKKLRAASLDELPELWNILKGEMSLVGPRPLLMQYLPL
               |||||||||||||||:|||| |||||||:||||||||||:||||||||||||||||||||
orf3ng         SMRDALDSDGIPLPDSERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRPLLMQYLPL
                       70        80        90       100       110       120

                      130       140       150       160       170       180
orf3-1.pep     YDNFQNRRHEMKPGITGWAQVNGRNALSWDEKFACDVWYIDHFSLCLDIKILLLTVKKVL
               |::||||||||||||||||||||||||||||||:||||| |:||: ||:|||:|||||||
orf3ng         YNKFQNRRHEMKPGITGWAQVNGRNALSWDEKFSCDVWYTDNFSFWLDMKILFLTVKKVL
                      130       140       150       160       170       180

                      190       200       210       220       230       240
orf3-1.pep     IKEGISAQGEATMPPFTGKRKLAVVGAGGHGKVVADLAAALGRYREIVFLDDRAQGSVNG
               ||||||||||||||||:|:|||||:||||||||||:|||||| | ||||||||:||||||
orf3ng         IKEGISAQGEATMPPFAGNRKLAVIGAGGHGKVVAELAAALGTYGEIVFLDDRTQGSVNG
                      190       200       210       220       230       240

                      250       260       270       280       290       300
orf3-1.pep     FSVIGTTLLLENSLSPEQYDVAVAVGNNRIRRQIAEKAAALGFALPVLVHPDATVSPSAT
               | ||||||||||||||||:|::||||||||||||:|:|||||| ||||:||||||||||
orf3ng         FPVIGTTLLLENSLSPEQFDITVAVGNNRIRRQITENAAALGFKLPVLIHPDATVSPSAI
                      250       260       270       280       290       300

                      310       320       330       340       350       360
orf3-1.pep     VGQGSVVMAKAVVQAGSVLKDGVIVNTAATVDHDCLLNAFVHISPGAHLSGNTHIGEESW
               :||||||||||||||||||||||||||||||||||||:|||||||||||||||:|||||
orf3ng         IGQGSVVMAKAVVQAGSVLKDGVIVNTAATVDHDCLLDAFVHISPGAHLSGNTRIGEESR
                      310       320       330       340       350       360

                      370       380       390       400       410
orf3-1.pep     IGTGACSRQQIRIGSRATIGAGAVVVRDVSDGMTVAGNPAKPLPRKNPETSTAX
               ||||||||||  :|| :| |||||:| |: |||||||||||||  |||:|:|||
orf3ng         IGTGACSRQQTTVGSGVTAGAGAVIVCDIPDGMTVAGNPAKPLTGKNPKTGTAX
                      370       380       390       400       410
```

In addition, ORF3ng shows significant homology with a hypothetical protein from *B.subtilis*:

```
gnl|PID|e238668 (Z71928) hypothetical protein [Bacillus subtilis]
>gi|1945702|gnl|PID|e313004 (Z94043) hypothetical protein [Bacillus subtilis]
>gi|2635938|gnl|PID|e1186113 (Z99121) similar to capsular polysaccharide
biosynthesis [Bacillus subtilis]Length = 202
 Score =  235 bits (594), Expect = 3e-61
 Identities = 114/195 (58%), Positives = 142/195 (72%)

Query: 5   VKRLFDIIASASGLIVLSPVFLVLIYLIRKNLGSPVFFIRERPGKDGKPFKMVKFRSMRD 64
            +KRLFD+ A+    L    S + L   I ++R  +GSPVFF + RPG  GKPF + KFR+M D
Sbjct: 3   LKRLFDLTAAIFLLCCTSVIILFTIAVVRLKIGSPVFFKQVRPGLHGKPFTLYKFRTMTD 62

Query: 65  ALDSDGIPLPDSERLTDFGKKLRATSLDELPELWNVLKGEMSLVGPRPLLMQYLPLYNKF 124
              DS G  LPD  RLT   G+ +R  S+DELP+L NVLKG++SLVGPRPLLM YLPLY +
Sbjct: 63  ERDSKGNLLPDEVRLTKTGRLIRKLSIDELPQLLNVLKGDLSLVGPRPLLMDYLPLYTEK 122

Query: 125 QNRRHEMKPGITGWAQVNGRNALSWDEKFSCDVWYTDNFSFWLDMKILFLTVKKVLIKEG 184
            Q RRHE+KPGITGWAQ+NGRNA+SW++KF  DVWY DN+SF+LD+KIL LTV+KVL+ EG
Sbjct: 123 QARRHEVKPGITGWAQINGRNAISWEKKFELDVWYVDNWSFFLDLKILCLTVRKVLVSEG 182

Query: 185 ISAQGEATMPPFAGN 199
            I      T   F G+
Sbjct: 183 IQQTNHVTAERFTGS 197
```

The hypothetical product of *yvfc* gene shows similarity to EXOY of *R.meliloti*, an exopolysaccharide production protein. Based on this and on the two predicted transmembrane regions in the homologous *N.gonorrhoeae* sequence, it is predicted that these proteins, or their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 4

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 19>:

```
  1  ..AACCATATGG CGATTGTCAT CGACGAATAC GGCGGCACAT CCGGCTTGGT
 51    CACCTTTGAA GACATCATCG AGCAAATCGT CGGCGAAATC GAAGACGAGT
101    TTGACGAAGA CGATAGCGCC GACAATATCC ATGCCGTTTC TTCAGACACG
151    TGGCGCATCC ATGCAGCTAC CGAAATCGAA GACATCAACA CCTTCTTCGG
201    CACGGAATAC AGCATCGAAG AAGCCGACAC CATT.GGCGG CCTGGTCATT
251    CAAGAGTTGG GACATCTGCC CGTGCGCGGC GAAAAAGTCC TTATCGGCGG
301    TTTGCAGTTC ACCGTCGCAC GCGCCGACAA CCGCCGCCTG CATACGCTGA
351    TGGCGACCCG CGTGAAGTAA GC........ .....ACCGC CGTTTCTGCA
401    CAGTTTAG
```

This corresponds to amino acid sequence <SEQ ID 20; ORF5>:

```
  1  ..NHMAIVIDEY GGTSGLVTFE DIIEQIVGEI EDEFDEDDSA DNIHAVSSDT
 51    WRIHAATEIE DINTFFGTEY SIEEADTIXR PGHSRVGTSA RARRKSPYRR
101    FAVHRRTRRQ PPPAYADGDP REVS....XR RFCTV*
```

Further sequence analysis revealed the complete DNA sequence to be <SEQ ID 21>:

```
  1  ATGGACGGCG CACAACCGAA AACGAATTTT TTTGAACGCC TGATTGCCCG
 51  ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTAAAC CTGCTTCGGC
101  AGGCGCACGA GCAGGAAGTT TTTGATGCGG ATACGCTTTT AAGATTGGAA
151  AAAGTCCTCG ATTTTTCCGA TTTGGAAGTG CGCGACGCGA TGATTACGCG
201  CAGCCGTATG AACGTTTTAA AAGAAAACGA CAGCATCGAG CGCATCACCG
251  CCTACGTTAT CGATACCGCC CATTCGCGCT TCCCCGTCAT CGGCGAAGAC
301  AAAGACGAAG TTTTGGGCAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351  GTTTAACCCC GAGCAGTTCC ACCTCAAATC CATTCTCCGC CCCGCCGTCT
401  TCGTCCCCGA AGGCAAATCG CTGACCGCCC TTTTAAAAGA GTTCCGCGAA
451  CAGCGCAACC ATATGGCGAT TGTCATCGAC GAATACGGCG GCACATCCGG
501  CTTGGTCACC TTTGAAGACA TCATCGAGCA AATCGTCGGC GAAATCGAAG
551  ACGAGTTTGA CGAAGACGAT AGCGCCGACA ATATCCATGC CGTTTCTTCC
601  GAACGCTGGC GCATCCATGC AGCTACCGAA ATCGAAGACA TCAACACCTT
651  CTTCGGCACG GAATACAGCA GCGAAGAAGC CGACACCATT CGGCCTGGTC
701  ATTCAAGAGT TGGGACATCT GCCCGTGCGC GGCGAAAAAG TCCTTATCGG
751  CGGTTTGCAG TTCACCGTCG CACGCGCCGA CAACCGCCGC CTGCATACGC
801  TGATGGCGAC CCGCGTGAAG TAAGCACCGC CGTTTCTGCA CAGTTTAGGA
851  TGACGGTACG GGCGTTTTCT GTTTCAATCC GCCCCATCCG CCAAACATAA
```

This corresponds to amino acid sequence <SEQ ID 22; ORF5-1>:

```
  1  MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLLRLE
 51  KVLDFSDLEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101  KDEVLGILHA KDLLKYMFNP EQFHLKSILR PAVFVPEGKS LTALLKEFRE
151  QRNHMAIVID EYGGTSGLVT FEDIIEQIVG EIEDEFDEDD SADNIHAVSS
201  ERWRIHAATE IEDINTFFGT EYSSEEADTI RPGHSRVGTS ARARRKSPYR
251  RFAVHRRTRR QPPPAYADGD PREVSTAVSA QFRMTVRAFS VSIRPIRQT*
```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 23 >:

```
  1  ATGGACGGCG CACAACCGAA AACAAATTTT TTNNAACGCC TGATTGCCCG
 51  ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTGACC CTGTTGCGCC
101  AAGCGCACGA ACAGGAAGTA TTTGATGCGG ATACGCTTTT AAGATTGGAA
151  AAAGTCCTCG ATTTTTCTGA TTTGGAAGTG CGCGACGCGA TGATTACGCG
201  CAGCCGTATG AACGTTTTAA AAGAAAACGA CAGCATCGAA CGCATCACCG
251  CCTACGTTAT CGATACCGCC CATTCGCGCT TCCCCGTCAT CGGTGAAGAC
301  AAAGACGAAG TTTTGGGTAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351  GTTCAACCCC GAGCAGTTCC ACCTCAAATC GATATTGCGC CCTGCCGTCT
```

```
401  TCGTCCCCGA AGGCAAATCG CTGACCGCCC TTTTAAAAGA GTTCCGCGAA
451  CAGCGCAACC ATATGGCAAT CGTCATCGAC GAATACGGCG GCACGTCGGG
501  TTTGGTAACT TTTGAAGACA TCATCGAGCA AATCGTCGGC GACATCGAAG
551  ATGAGTTTGA CGAAGACGAA AGCGCGGACA ACATCCACGC CGTTTCCGCC
601  GAACGCTGGC GCATCCACGC GGCTACCGAA ATCGAAGACA TCAACGCCTT
651  TTTCGGCACG GAATACAGCA GCGAAGAAGC CGACACCATC GGCGGCCNTG
701  GTCATTCAGG AATTGGNACA CCTGCCCGTG CGCGGCGAAA AAGTCNTTAT
751  CGGCGNNTTG CANTTCACNG TCGCCNGCGC NGACAACCGC CGCCTGCATA
801  CGCTGATGGC GACCCGCGTG AAGTAAGCTC CGCCGTTTCT GTACAGTTTA
851  GGATGACGGT ACGGGCGTTT TCTGTTTCAA TCCGCCCCAT CCGCCANACA
901  TAA
```

This encodes a protein having amino acid sequence <SEQ ID 24; ORF5a>:

```
  1  MDGAQPKTNF XXRLIARLAR EPDSAEDVLT LLRQAHEQEV FDADTLLRLE
 51  KVLDFSDLEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101  KDEVLGILHA KDLLKYMFNP EQFHLKSILR PAVFVPEGKS LTALLKEFRE
151  QRNHMAIVID EYGGTSGLVT FEDIIEQIVG DIEDEFDEDE SADNIHAVSA
201  ERWRIHAATE IEDINAFFGT EYSSEEADTI GGXGHSGIGT PARARRKSXY
251  RRXAXHXRXR XQPPPAYADG DPREVSSAVS VQFRMTVRAF SVSIRPIRXT
301  *
```

The originally-identified partial strain B sequence (ORF5) shows 54.7% identity over a 124aa overlap with ORF5a:

```
                                                        10        20        30
orf5.pep                                        NHMAIVIDEYGGTSGLVTFEDIIEQIVGEI
                                                ||||||||||||||||||||||||||||:|
orf5a             FHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVGDI
                       130       140       150       160       170       180

                          40        50        60        70        80        90
orf5.pep          EDEFDEDDSADNIHAVSSDTWRIHAATEIEDINTFFGTEYSIEEADTIXRPGHSRVGTSA
                  |||||||:|||||||||:: |||||||||||||:||||||| ||||||   ||| :|| |
orf5a             EDEFDEDESADNIHAVSAERWRIHAATEIEDINAFFGTEYSSEEADTIGGXGHSGIGTPA
                       190       200       210       220       230       240

                         100       110       120       130
orf5.pep          RARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSXXXXXRRFCTV
                  |||||| ||| | | |:| |||||||||||||||
orf5a             RARRKSXYRRXAXHXRXRXQPPPAYADGDPREVSSAVSVQFRMTVRAFSVSIRPIRXTX
                       250       260       270       280       290       300
```

The complete strain B sequence (ORF5-1) and ORF5a show 92.7% identity in 300 aa overlap:

```
                          10        20        30        40        50        60
orf5a.pep         MDGAQPKTNFXXRLIARLAREPDSAEDVLTLLRQAHEQEVFDADTLLRLEKVLDFSDLEV
                  ||||||||||  |||||||||||||||||:||||||||||||||||||||||||||||||
orf5-1            MDGAQPKTNFFERLIARLAREPDSAEDVLNLLRQAHEQEVFDADTLLRLEKVLDFSDLEV
                          10        20        30        40        50        60

                          70        80        90       100       110       120
orf5a.pep         RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFNP
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf5-1            RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFNP
                          70        80        90       100       110       120

                         130       140       150       160       170       180
orf5a.pep         EQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVG
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf5-1            EQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVG
                         130       140       150       160       170       180

                         190       200       210       220       230       240
orf5a.pep         DIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAFFGTEYSSEEADTIGGXGHSGIGT
                  :||||||||:|||||||||:|||||||||||||||:||||||||||||||   ||| :||
orf5-1            EIEDEFDEDDSADNIHAVSSERWRIHAATEIEDINTFFGTEYSSEEADTIRP-GHSRVGT
                         190       200       210       220       230

                         250       260       270       280       290       300
```

```
orf5a.pep    PARARRKSXYRRXAXHXRXRXQPPPAYADGDPREVSSAVSVQFRMTVRAFSVSIRPIRXT
              ||||||| ||| | | |:| |||||||||||||||:|||:|||||||||||||||| |
orf5-1       SARARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSTAVSAQFRMTVRAFSVSIRPIRQT
           240       250       260       270       280       290
```

Further work identified the a partial DNA sequence in *N.gonorrhoeae* <SEQ ID 25> which encodes a protein having amino acid sequence <SEQ ID 26; ORF5ng>:

```
  1 MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLTRLE
 51 KVLDFAELEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101 KDEVLGILHA KDLLKYMFNP EQFHLKSVLR PAVFVPEGKS LTALLKEFRE
151 QRNHMAIVID EYGGTSGLVT FEDIIEQIVG DIEDEFDEDE SADDIHSVSA
201 ERWRIHAATE IEDINAFFGT EYGSEEADTI RRLGHSGIGT PARARRKSPY
251 RRFAVHRRPR RQPPPAHADG DPREVSRACP HRRFCTV*
```

Further analysis revealed the complete gonococcal nucleotide sequence <SEQ ID 27> to be:

```
  1 ATGGACGGCG CACAACCGAA AACAAATTTT TTTGAACGCC TGATTGCCCG
 51 ACTCGCCCGC GAACCCGATT CCGCCGAAGA CGTATTAAAC CTGCTTCGGC
101 AGGCGCACGA ACAGGAAGTT TTTGATGCCG ACACACTGAC CCGGCTGGAA
151 AAAGTATTGG ACTTTGCCGA GCTGGAAGTG CGCGATGCGA TGATTACGCG
201 CAGCCGCATG AACGTATTGA AAGAAAACGA CAGCATCGAA CGCATCACCG
251 CCTACGTCAT CGATACCGCC CATTCGCGCT TCCCCGTCAT CGGCGAAGAC
301 AAAGACGAAG TTTTGGGCAT TTTGCACGCC AAAGACCTGC TCAAATATAT
351 GTTCAACCCC GAGCAGTTCC ACCTGAAATC CGTCTTGCGC CCTGCCGTTT
401 TCGTGCCCGA AGGCAAATCT TTGACCGCCC TTTTAAAAGA GTTCCGCGAA
451 CAGCGCAACC ATATGGCAAT CGTCATCGAC GAATACGGCG GCACGTCGGG
501 TTTGGTCACC TTTGAAGACA TCATCGAGCA AATCGTCGGT GACATCGAAG
551 ACGAGTTTGA CGAAGACGAA AGCGccgacg acatCCACTC cgTTTccgCC
601 GAACGCTGGC GCATCCacgc ggctaCCGAA ATCGAAGaca TCAACGCCTT
651 TTTCGGTACG GAatacggca gcgaagaagc cgacaccatc cggcggctTG
701 GTCATTCAGG AATTGGGACA CCTGCCCGTG CGCGGCGAAA AAGTCCTTAt
751 cggcgGTTTG Cagttcaccg tCGCCCGCGC CGACAACCGC CGCCTGCACA
801 CGCTGATGGC GACCCGCGTG AAGTAAGCAG AGCCTGCCcg AccgccgttT
851 CTGCacAGTT TAGGatgACG gtaCGGTCGT TTTCTGTTTC AATCCGCCCC
901 ATCCGCCAAA CATAA
```

This encodes a protein having amino acid sequence <SEQ ID 28; ORF5ng-1>:

```
  1 MDGAQPKTNF FERLIARLAR EPDSAEDVLN LLRQAHEQEV FDADTLTRLE
 51 KVLDFAELEV RDAMITRSRM NVLKENDSIE RITAYVIDTA HSRFPVIGED
101 KDEVLGILHA KDLLKYMFNP EQFHLKSVLR PAVFVPEGKS LTALLKEFRE
151 QRNHMAIVID EYGGTSGLVT FEDIIEQIVG DIEDEFDEDE SADDIHSVSA
201 ERWRIHAATE IEDINAFFGT EYGSEEADTI RRLGHSGIGT PARARRKSPY
251 RRFAVHRRPR RQPPPAHADG DPREVSRACP TAVSAQFRMT VRSFSVSIRP
301 IRQT*
```

The originally-identified partial strain B sequence (ORF5) shows 83.1% identity over a 135aa overlap with the partial gonococcal sequence (ORF5ng):

```
orf5                                     NHMAIVIDEYGGTSGLVTFEDIIEQIVGEI    30
                                         ||||||||||||||||||||||||||||:|
orf5ng       FHLKSVLRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVGDI   182

orf5         EDEFDEDDSADNIHAVSSDTWRIHAATEIEDINTFFGTEYSIEEADTIXRPGHSRVGTSA    90
             |||||||:|||:||:||::  ||||||||||||:||||||:  |||||| | ||| :|| |
orf5ng       EDEFDEDESADDIHSVSAERWRIHAATEIEDINAFFGTEYGSEEADTIRRLGHSGIGTPA   242

orf5         RARRKSPYRRFAVHRRTRRQPPPAYADGDPREVSX----RRFCTV  131
             ||||||||||||||| ||||||:|||||||||       ||||||
orf5ng       RARRKSPYRRFAVHRRPRRQPPPAHADGDPREVSRACPHRRFCTV  287
```

The complete strain B and gonococcal sequences (ORF5-1 & ORF5ng-1) show 92.4% identity in 304 aa overlap:

```
                     10        20        30        40        50        60
orf5ng-1.pep MDGAQPKTNFFERLIARLAREPDSAEDVLNLLRQAHEQEVFDADTLTRLEKVLDFAELEV
```

```
               |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||::|||
orf5-1         MDGAQPKTNFFERLIARLAREPDSAEDVLNLLRQAHEQEVFDADTLLRLEKVLDFSDLEV
                       10        20        30        40        50        60

                       70        80        90       100       110       120
orf5ng-1.pep   RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFNP
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf5-1         RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYMFNP
                       70        80        90       100       110       120

                      130       140       150       160       170       180
orf5ng-1.pep   EQFHLKSVLRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVG
               ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
orf5-1         EQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIVG
                      130       140       150       160       170       180

                      190       200       210       220       230       240
orf5ng-1.pep   DIEDEFDEDESADDIHSVSAERWRIHAATEIEDINAFFGTEYGSEEADTIRRLGHSGIGT
               :|||||||:|||:||:||:|||||||||||||||:||||||:|||||||| ||| :||
orf5-1         EIEDEFDEDDSADNIHAVSSERWRIHAATEIEDINTFFGTEYSSEEADTIRP-GHSRVGT
                      190       200       210       220       230

                      250       260       270       280       290       300
orf5ng-1.pep   PARARRKSPYRRFAVHRRPRRQPPPAHADGDPREVSRACPTAVSAQFRMTVRSFSVSIRP
               ||||||||||||||||| |||||||:||||||||     ||||||||||||:|||||||
orf5-1         SARARRKSPYRRFAVHRRTRRQPPPAYADGDPREVS----TAVSAQFRMTVRAFSVSIRP
             240       250       260       270           280       290

orf5ng-1.pep   IRQTX
               |||||
orf5-1         IRQTX
                 300
```

Computer analysis of these amino acid sequences indicates a putative leader sequence, and identified the following homologies:

<u>Homology with hemolysin homolog TlyC (accession U32716) of *H.influenzae*</u>

ORF5 and TlyC proteins show 58% aa identity in 77 aa overlap (BLASTp).

```
ORF5    2   HMAIVIDEYGGTSGLVTFEDIIEQIVGEIEDEFDEDDSADNIHAVSSDTWRIHAATEIED 61
            HMAIV+DE+G   SGLVT EDI+EQIVG+IEDEFDE++ AD I   +S   T+ + A T+I+D
TlyC    166 HMAIVVDEFGAVSGLVTIEDILEQIVGDIEDEFDEEEIAD-IRQLSRHTYAVRALTDIDD 224

ORF5    62  INTFFGTEYSIEEADTI 78
            N  F T++  EE DTI
TlyC    225 FNAQFNTDFDDEEVDTI 241
```

ORF5ng-1 also shows significant homology with TlyC:

```
SCORES      Init1:   301 Initn:   419 Opt:   668
Smith-Waterman score: 668;    45.9% identity in 242 aa overlap

                                  10        20        30        40        50
orf5ng-1.pep         MDGAQPKTNFFERLIARLAR-EPDSAEDVLNLLRQAHEQEVFDADTLTRLEK
                            |   ||: |::|: :  |   : |::::::|::::::::| :|    :|
tlyc_haein     MNDEQQNSNQSENTKKPFFQSLFGRFFQGELKNREELVEVIRDSEQNDLIDQNTREMIEG
                       10        20        30        40        50        60

                       60        70        80        90          100       109
orf5ng-1.pep   VLDFAELEVRDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGE--DKDEVLGILH
               |:::|||:||| || ||::  ::::::::      :|::||||||||||::   |:|:::||||
tlyc_haein     VMEIAELRVRDIMIPRSQIIFIEDQQDLNTCLNTIIESAHSRFPVIADADDRDNIVGILH
                       70        80        90       100       110       120

             110       120       130       140       150       160
orf5ng-1.pep AKDLLKYMF-NPEQFHLKSVLRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGL
               ||||||::   :  |  |  |:|:|||:|:|||:|  :   :||:||  :|  |||||:||:|::|||
tlyc_haein     AKDLLKFLREDAEVFDLSSLLRPVVIVPESKRVDRMLKDFRSERFHMAIVVDEFGAVSGL
```

```
                        130       140       150       160       170       180

                170       180       190       200       210       220
orf5ng-1.pep VTFEDIIEQIVGDIEDEFDEDESADDIHSVSAERWRIHAATEIEDINAFFGTEYGSEEAD
             ||:|||:|||||||||||||||:| || |:::| : : ::| |:|:|:|| |:|:: :||:|
tlyc_haein   VTIEDILEQIVGDIEDEFDEEEIAD-IRQLSRHTYAVRALTDIDDFNAQFNTDFDDEEVD
                     190       200       210       220       230

                230       240       250       260       270       280
orf5ng-1.pep TIRRLGHSGIG-TPARARRKSPYRRFAVHRRPRRQPPPAHADGDPREVSRACPTAVSAQF
             ||  |   : :|   | |:
tlyc_haein   TIGGLIMQTFGYLPKRGEEIILKNLQFKVTSADSRRLIQLRVTVPDEHLAEMNNVDEKSE
           240       250       260       270       280       290
```

<u>Homology with a hypothetical secreted protein from *E.coli*:</u>

ORF5a shows homology to a hypothetical secreted protein from *E.coli*:

sp|P77392|YBEX_ECOLI HYPOTHETICAL 33.3 KD PROTEIN IN CUTE-ASNB INTERGENIC REGION >gi|1778577 (U82598) similar to *H. influenzae* [*Escherichia coli*] >gi|1786879 (AE000170) f292; This 292 aa ORF is 23% identical (9 gaps) to 272 residues of an approx. 440 aa protein YTFL_HAEIN SW: P44717 [*Escherichia coli*] Length = 292

```
 Score =  212 bits (533), Expect = 3e-54
 Identities = 112/230 (48%), Positives = 149/230 (64%), Gaps = 3/230 (1%)

Query: 2   DGAQPKTNFXXRLIARLAR-EPDSAEDVLTLLRQAHEQEVFDADTLLRLEKVLDFSDLEV 60
            D    K  F   L+++L    EP + +++L L+R + + ++ D DT    LE V+D +D  V
Sbjct: 10  DTISNKKGFFSLLLSQLFHGEPKNRDELLALIRDSGQNDLIDEDTRDMLEGVMDIADQRV 69

Query: 61  RDAMITRSRMNVLKENDSIERITAYVIDTAHSRFPVIGEDKDEVLGILHAKDLLKYM-FN 119
            RD MI RS+M  LK N +++      +I++AHSRFPVI EDKD + GIL AKDLL +M  +
Sbjct: 70  RDIMIPRSQMITLKRNQTLDECLDVIIESAHSRFPVISEDKDHIEGILMAKDLLPFMRSD 129

Query: 120 PEQFHLKSILRPAVFVPEGKSLTALLKEFREQRNHMAIVIDEYGGTSGLVTFEDIIEQIV 179
              E F +   +LR AV VPE K +   +LKEFR QR HMAIVIDE+GG SGLVT EDI+E IV
Sbjct: 130 AEAFSMDKVLRQAVVVPESKRVDRMLKEFRSQRYHMAIVIDEFGGVSGLVTIEDILELIV 189

Query: 180 GDIEDEFDEDESADNIHAVSAERWRIHAATEIEDINAFFGTEYSSEEADT 229
            G+IEDE+DE++  D    +S   W + A    IED N  FGT +S EE DT
Sbjct: 190 GEIEDEYDEEDDID-FRQLSRHTWTVRALASIEDFNEAFGTHFSDEEVDT 238
```

Based on this analysis, including the amino acid homology to the TlyC hemolysin-homologue from *H. influenzae* (hemolysins are secreted proteins), it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae* are secreted and could thus be useful antigens for vaccines or diagnostics.

ORF5-1 (30.7kDa) was cloned in the pGex vector and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 2A shows the results of affinity purification of the GST-fusion protein. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot analysis (Figure 1B). These experiments confirm that ORF5-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 5

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 29>:

```
 1  ATGCGCGGCG GCAGGCCGGA TTCCGTTACC GTGCAGATTA TCGAAGGTTC
51  GCGTTTTTCG CATATGAGGA AAGTCATCGA CGCAACGCCC GACATCGGAC
```

```
101  ACGACACCAA AGGCTGGAGC AATGAAAAAC TGATGGCGGA AGTTGCGCCC
151  GATGCCTTCA GCGGCAATCC TGAAgGGCAG TTTTTCCCCG ACAGCTACGA
201  AATCGATGCG GGCGGCAGTG ATTTGCAGAT TTACCAAACC GCCTACAAgG
251  GCGATGCAAC GCCGCCTGAA TGAgGGCATG GGAAAGCAGG CAGGACGGGC
301  TGCCTTATAA AAACCCTTAT GAAATGCTGA TTATGGCGAr CCTGGTCGAA
351  AAGGAAACAG GGCATGAAGC CGAsCsCGAC CATGTcGCTT CCGTCTTCGT
401  CAACCGCCTG AAAATCGGTA TGCGCCTGCA AACCgAssCG TCCGTGATTT
451  ACGGCATGGG TGCGGCATAC AAGGGCAAAA TCCGTAAAGC CGACCTGCGC
501  CGCGACACGC CGTACAACAC CTACACGCGC GGCGGTCTGC CGCCAACCCC
551  GATTGCGCTG CCC..
```

This corresponds to the amino acid sequence <SEQ ID 30; ORF7>:

```
  1  MRGGRPDSVT VQIIEGSRFS HMRKVIDATP DIGHDTKGWS NEKLMAEVAP
 51  DAFSGNPEGQ FFPDSYEIDA GGSDLQIYQT AYKAMQRRLN EAWESRQDGL
101  PYKNPYEMLI MAXLVEKETG HEAXXDHVAS VFVNRLKIGM RLQTXXSVIY
151  GMGAAYKGKI RKADLRRDTP YNTYTRGGLP PTPIALP..
```

Further sequence analysis revealed the complete DNA sequence <SEQ ID 31>:

```
  1  ATGTTGAGAA AATTGTTGAA ATGGTCTGCC GTTTTTTTGA CCGTGTCGGC
 51  AGCCGTTTTC GCCGCGCTGC TTTTTGTTCC TAAGGATAAC GGCAGGGCAT
101  ACCGAATCAA AATTGCCAAA AACCAGGGTA TTTCGTCGGT CGGCAGGAAA
151  CTTGCCGAAG ACCGCATCGT GTTCAGCAGG CATGTTTTGA CGGCGGCGGC
201  CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGGACG TACAGATTGC
251  CTTCGGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CGGCGGCAGG
301  CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTCGCGTT TTTCGCATAT
351  GAGGAAAGTC ATCGACGCAA CGCCCGACAT CGGACACGAC ACCAAAGGCT
401  GGAGCAATGA AAAACTGATG GCGGAAGTTG CGCCCGATGC CTTCAGCGGC
451  AATCCTGAAG GGCAGTTTTT CCCCGACAGC TACGAAATCG ATGCGGGCGG
501  CAGTGATTTG CAGATTTACC AAACCGCCTA CAAGGCGATG CAACGCCGCC
551  TGAATGAGGC ATGGGAAAGC AGGCAGGACG GGCTGCCTTA TAAAAACCCT
601  TATGAAATGC TGATTATGGC GAGCCTGGTC GAAAAGGAAA CAGGGCATGA
651  AGCCGACCGC GACCATGTCG CTTCCGTCTT CGTCAACCGC CTGAAAATCG
701  GTATGCGCCT GCAAACCGAC CCGTCCGTGA TTTACGGCAT GGGTGCGGCA
751  TACAAGGGCA AAATCCGTAA AGCCGACCTG CGCCGCGACA CGCCGTACAA
801  CACCTACACG CGCGGCGGTC TGCCGCCAAC CCCGATTGCG CTGCCCGGCA
851  AGGCGGCACT CGATGCCGCC GCCCATCCGT CCGGCGAAAA ATACCTGTAT
901  TTCGTGTCCA AAATGGACGG CACGGGCTTG AGCCAGTTCA GCCATGATTT
951  GACCGAACAC AATGCCGCCG TCCGCAAATA TATTTTGAAA AAATAA
```

This corresponds to the amino acid sequence <SEQ ID 32; ORF7-1>:

```
  1  MLRKLLKWSA VFLTVSAAVF AALLFVPKDN GRAYRIKIAK NQGISSVGRK
 51  LAEDRIVFSR HVLTAAAYVL GVHNRLHTGT YRLPSEVSAW DILQKMRGGR
101  PDSVTVQIIE GSRFSHMRKV IDATPDIGHD TKGWSNEKLM AEVAPDAFSG
151  NPEGQFFPDS YEIDAGGSDL QIYQTAYKAM QRRLNEAWES RQDGLPYKNP
201  YEMLIMASLV EKETGHEADR DHVASVFVNR LKIGMRLQTD PSVIYGMGAA
251  YKGKIRKADL RRDTPYNTYT RGGLPPTPIA LPGKAALDAA AHPSGEKYLY
301  FVSKMDGTGL SQFSHDLTEH NAAVRKYILK K*
```

Computer analysis of this amino acid sequence gave the following results:

<u>Homology with hypothetical protein encoded by *yceg* gene (accession P44270) of *H.influenzae*</u>

ORF7 and yceg proteins show 44% aa identity in 192 aa overlap:

```
ORF7   1   MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMA-----EVAPDAFSG 55
           +  G+     V+ IEG  F   RK ++  P +    K  SNE++ A     ++  +
yceg   102 LNSGKEVQFNVKWIEGKTFKDWRKDLENAPHLVQTLKDKSNEEIFALLDLPDIGQNLELK 161

ORF7   56  NPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLNEAWESRQDGLPYKNPYEMLIMAXLV 115
           N EG  +PD+Y      +DL++ + + + M++ LN+AW  R + LP  NPYEMLI+A +V
yceg   162 NVEGWLYPDTYNYTPKSTDLELLKRSAERMKKALNKAWNERDEDLPLANPYEMLILASIV 221

ORF7   116 EKETGHEAXXDHVASVFVNRLKIGMRLQTXXSVIYGMGAAYKGKIRKADLRRDTPYNTYT 175
           EKETG       VASVF+NRLK  M+LQT  +VIYGMG  Y G IRK DL    TPYNTY
yceg   222 EKETGIANERAKVASVFINRLKAKMKLQTDPTVIYGMGENYNGNIRKKDLETKTPYNTYV 281
```

```
ORF7   176 RGGLPPTPIALP 187
             GLPPTPIA+P
yceg   282 IDGLPPTPIAMP 293
```

The complete length YCEG protein has sequence:

```
  1 MKKFLIAILL LILILAGVAS FSYYKMTEFV KTPVNVQADE LLTIERGTTS
 51 SKLATLFEQE KLIADGKLLP YLLKLKPELN KIKAGTYSLE NVKTVQDLLD
101 LLNSGKEVQF NVKWIEGKTF KDWRKDLENA PHLVQTLKDK SNEEIFALLD
151 LPDIGQNLEL KNVEGWLYPD TYNYTPKSTD LELLKRSAER MKKALNKAWN
201 ERDEDLPLAN PYEMLILASI VEKETGIANE RAKVASVFIN RLKAKMKLQT
251 DPTVIYGMGE NYNGNIRKKD LETKTPYNTY VIDGLPPTPI AMPSESSLQA
301 VANPEKTDFY YFVADGSGGH KFTRNLNEHN KAVQEYLRWY RSQKNAK
```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF7 shows 95.2% identity over a 187aa overlap with an ORF (ORF7a) from strain A of *N. meningitidis:*

```
                                                  10        20        30
orf7.pep                                  MRGGRPDSVTVQIIEGSRFSHMRKVIDATP
                                          ||||||||||||||||||||||||||||||
orf7a       AAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKVIDATP
                  70        80        90       100       110       120

                    40        50        60        70        80        90
orf7.pep    DIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLN
            || ||||||||||||||||||||||||||||||||||||||||||:||| ||||||||||
orf7a       DIEHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAMQRRLN
                   130       140       150       160       170       180

                   100       110       120       130       140       150
orf7.pep    EAWESRQDGLPYKNPYEMLIMAXLVEKETGHEAXXDHVASVFVNRLKIGMRLQTXXSVIY
            |||||||||||||||||||||| |:||||||||  |||||||||||||||||||  ||||
orf7a       EAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFVNRLKIGMRLQTDPSVIY
                   190       200       210       220       230       240

                   160       170       180
orf7.pep    GMGAAYKGKIRKADLRRDTPYNTYTRGGLPPTPIALP
            |||||||||||||||||||||||||||||||||||||
orf7a       GMGAAYKGKIRKADLRRDTPYNTYTRGGLPPTPIALPGKAALDAAAHPSGEKYLYFVSKM
                   250       260       270       280       290       300

orf7a       DGTGLSQFSHDLTEHNAAVRKYILKKX
                   310       320       330
```

The complete length ORF7a nucleotide sequence <SEQ ID 33> is:

```
  1 ATGTTGAGAA AATTGTTGAA ATGGTCTGCC GTTTTTTTGA CCGTATCGGC
 51 AGCCGTTTTC GCCGCGCTGC TTTTCGTCCC TAAAGACAAC GGCAGGGCAT
101 ACAGGATTAA AATTGCCAAA AACCAGGGTA TTTCGTCGGT CGGCAGGAAA
151 CTTGCCGAAG ACCGCATCGT GTTCAGCAGG CATGTTTTGA CGGCGGCGGC
201 CTACGTTTTG GGTGTGCACA ACAGGCTGCA TACGGGGACG TACAGACTGC
251 CTTCGGAAGT GTCTGCTTGG GATATCTTGC AGAAAATGCG CGGCGGCAGG
301 CCGGATTCCG TTACCGTGCA GATTATCGAA GGTTCGCGTT TTTCGCATAT
351 GAGGAAAGTC ATCGACGCAA CGCCCGACAT CGAACACGAC ACCAAAGGCT
401 GGAGCAATGA AAAACTGATG GCGGAAGTTG CCCCTGATGC CTTCAGCGGC
451 AATCCTGAAG GGCAGTTTTT CCCCGACAGC TACGAAATCG ATGCGGGCGG
501 CAGCGATTTA CGGATTTACC AAATCGCCTA CAAGGCGATG CAACGCCGAC
551 TGAATGAGGC ATGGGAAAGC AGGCAGGACG GGCTGCCTTA TAAAAACCCT
601 TATGAAATGC TGATTATGGC GAGCCTGATC GAAAAGGAAA CAGGGCATGA
651 AGCCGACCGC GACCATGTCG CTTCCGTCTT CGTCAACCGC CTGAAAATCG
701 GTATGCGCCT GCAAACCGAC CCGTCCGTGA TTTACGGCAT GGGTGCGGCA
751 TACAAGGGCA AAATCCGTAA AGCCGACCTG CGCCGCGACA CGCCGTACAA
801 CACCTACACG CGCGGCGGTC TGCCGCCAAC CCCGATCGCG CTGCCCGGCA
851 AGGCGGCACT CGATGCCGCC GCCCATCCGT CCGGTGAAAA ATACCTGTAT
901 TTCGTGTCCA AAATGGACGG TACGGGCTTG AGCCAGTTCA GCCATGATTT
951 GACCGAACAC AACGCCGCCG TTCGCAAATA TATTTGAAA AAATAA
```

This is predicted to encode a protein having amino acid sequence <SEQ ID 34>:

```
  1  MLRKLLKWSA VFLTVSAAVF AALLFVPKDN GRAYRIKIAK NQGISSVGRK
 51  LAEDRIVFSR HVLTAAAYVL GVHNRLHTGT YRLPSEVSAW DILQKMRGGR
101  PDSVTVQIIE GSRFSHMRKV IDATPDIEHD TKGWSNEKLM AEVAPDAFSG
151  NPEGQFFPDS YEIDAGGSDL RIYQIAYKAM QRRLNEAWES RQDGLPYKNP
201  YEMLIMASLI EKETGHEADR DHVASVFVNR LKIGMRLQTD PSVIYGMGAA
251  YKGKIRKADL RRDTPYNTYT RGGLPPTPIA LPGKAALDAA AHPSGEKYLY
301  FVSKMDGTGL SQFSHDLTEH NAAVRKYILK K*
```

A leader peptide is underlined.

ORF7a and ORF7-1 show 98.8% identity in 331 aa overlap:

```
                    10        20        30        40        50        60
orf7a.pep   MLRKLLKWSAVFLTVSAAVFAALLFVPKDNGRAYRIKIAKNQGISSVGRKLAEDRIVFSR
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf7-1      MLRKLLKWSAVFLTVSAAVFAALLFVPKDNGRAYRIKIAKNQGISSVGRKLAEDRIVFSR
                    10        20        30        40        50        60

                    70        80        90       100       110       120
orf7a.pep   HVLTAAAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKV
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf7-1      HVLTAAAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKV
                    70        80        90       100       110       120

                   130       140       150       160       170       180
orf7a.pep   IDATPDIEHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLRIYQIAYKAM
            ||||||| ||||||||||||||||||||||||||||||||||||||||||:||| |||||
orf7-1      IDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAM
                   130       140       150       160       170       180

                   190       200       210       220       230       240
orf7a.pep   QRRLNEAWESRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFVNRLKIGMRLQTD
            |||||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
orf7-1      QRRLNEAWESRQDGLPYKNPYEMLIMASLVEKETGHEADRDHVASVFVNRLKIGMRLQTD
                   190       200       210       220       230       240

                   250       260       270       280       290       300
orf7a.pep   PSVIYGMGAAYKGKIRKADLRRDTPYNTYTRGGLPPTPIALPGKAALDAAAHPSGEKYLY
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf7-1      PSVIYGMGAAYKGKIRKADLRRDTPYNTYTRGGLPPTPIALPGKAALDAAAHPSGEKYLY
                   250       260       270       280       290       300

                   310       320       330
orf7a.pep   FVSKMDGTGLSQFSHDLTEHNAAVRKYILKKX
            |||||||||||||||||||||||||||||||
orf7-1      FVSKMDGTGLSQFSHDLTEHNAAVRKYILKKX
                   310       320       330
```

<u>Homology with a predicted ORF from *N.gonorrhoeae*</u>

ORF7 shows 94.7% identity over a 187aa overlap with a predicted ORF (ORF7.ng) from *N. gonorrhoeae*:

```
orf7      MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQ   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf7ng    MRGGRPDSVTVQIIEGSRFSHMRKVIDATPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQ   60

orf7      FFPDSYEIDAGGSDLQIYQTAYKAMQRRLNEAWESRQDGLPYKNPYEMLIMAXLVEKETG  120
          ||||||||||||||||||||||||||||||||| :||||||||||||||||| |:|||||
orf7ng    FFPDSYEIDAGGSDLQIYQTAYKAMQRRLNEAWAGRQDGLPYKNPYEMLIMASLIEKETG  120

orf7      HEAXXDHVASVFVNRLKIGMRLQTXXSVIYGMGAAYKGKIRKADLRRDTPYNTYTRGGLP  180
          |||  |||||||||||||||||||  ||||||||||||||||||||||||||||| ||||
orf7ng    HEADRDHVASVFVNRLKIGMRLQTDPSVIYGMGAAYKGKIRKADLRRDTPYNTYTGGGLP  180

orf7      PTPIALP                                                      187
```

```
                 || ||||
orf7ng        PTRIALPGKAAMDAAAHPSGEKYLYFVSKMDGTGLSQFSHDLTEHNAAVRKYILKK  236
```

An ORF7ng nucleotide sequence <SEQ ID 35> is predicted to encode a protein having amino acid sequence <SEQ ID 36>:

```
  1  MRGGRPDSVT VQIIEGSRFS HMRKVIDATP DIGHDTKGWS NEKLMAEVAP
 51  DAFSGNPEGQ FFPDSYEIDA GGSDLQIYQT AYKAMQRRLN EAWAGRQDGL
101  PYKNPYEMLI MASLIEKETG HEADRDHVAS VFVNRLKIGM RLQTDPSVIY
151  GMGAAYKGKI RKADLRRDTP YNTYTGGGLP PTRIALPGKA AMDAAAHPSG
201  EKYLYFVSKM DGTGLSQFSH DLTEHNAAVR KYILKK*
```

Further sequence analysis revealed a partial DNA sequence of ORF7ng <SEQ ID 37>:

```
  1  ..taccgaatca AGATTGCCAA AAATCAGGGT ATTTCGTCGG TCGGCAGGAA
 51    ACTTGCcgaA GACCGCATCG TGTTCAGCAG GCATGTTTTG ACAGCGGCGG
101    CCTACGTTTT GGGTGTGCAC AACAGGCTGC ATACGGGGAC gTACAGATTG
151    CCTTCGGAAG TGTCTGCTTG GGATATCTTG CAGAAAATGC GCGGCGGCAG
201    GCCGGATTCC GTTACCGTGC AGATTATCGA AGGTTCGCGT TTTTCGCATA
251    TGAGGAAAGT CATCGACGCA ACGCCCGACA TCGGACACGA CACCAAAGGC
301    TGGAGCAATG AAAAACTGAT GGCGGAAGTT GCGCCCGATG CCTTCAGCGG
351    CAATCCTGAA GGGCAGTTTT TTCCCGACAG CTACGAAATC GATGCGGGCG
401    GCAGCGATTT GCAGATTTAC CAAACCGCCT ACAAGGCGAT GCAACGCCGC
451    CTGAACGAGG CATGGGCAGG CAGGCAGGAC GGGCTGCCTT ATAAAAACCC
501    TTATGAAATG CTGATTATGG CGAGCCTGAT CGAAAAGGAA ACGGGGCATG
551    AGGCCGACCG CGACCATGTC GCTTCCGTCT TCGTCAACCG CCTGAAAATC
601    GGTATGCGCC TGCAAACCGA CCCGTCCGTG ATTTACGGCA TGGGTGCGGC
651    ATACAAGGGC AAAATCCGTA AAGCCGACCT GCGCCGCGAC ACGCCGTACA
701    aCAccTAtac gggcgggggc ttgccgccaa cccggattgc gctgcccggC
751    Aaggcggcaa tggatgccgc cgcccacccg tccggcgaAa aatacctgTa
801    tttcgtgtcC AAAATGGACG GCACGGGCTT GAGCCAGTTC AGCCATGATT
851    TGACCGAACA CAACGCCGCc gTcCGCAAAT ATATTTTGAA AAAATAA
```

This corresponds to the amino acid sequence <SEQ ID 38; ORF7ng-1>:

```
  1  ..YRIKIAKNQG ISSVGRKLAE DRIVFSRHVL TAAAYVLGVH NRLHTGTYRL
 51    PSEVSAWDIL QKMRGGRPDS VTVQIIEGSR FSHMRKVIDA TPDIGHDTKG
101    WSNEKLMAEV APDAFSGNPE GQFFPDSYEI DAGGSDLQIY QTAYKAMQRR
151    LNEAWAGRQD GLPYKNPYEM LIMASLIEKE TGHEADRDHV ASVFVNRLKI
201    GMRLQTDPSV IYGMGAAYKG KIRKADLRRD TPYNTYTGGG LPPTRIALPG
251    KAAMDAAAHP SGEKYLYFVS KMDGTGLSQF SHDLTEHNAA VRKYILKK*
```

ORF7ng-1 and ORF7-1 show 98.0% identity in 298 aa overlap:

```
                     10        20        30        40        50        60
orf7-1.pep    KLLKWSAVFLTVSAAVFAALLFVPKDNGRAYRIKIAKNQGISSVGRKLAEDRIVFSRHVL
                                              ||||||||||||||||||||||||||||||
orf7ng-1                                      YRIKIAKNQGISSVGRKLAEDRIVFSRHVL
                                                      10        20        30

                     70        80        90       100       110       120
orf7-1.pep    TAAAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKVIDA
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf7ng-1      TAAAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPDSVTVQIIEGSRFSHMRKVIDA
                      40        50        60        70        80        90

                    130       140       150       160       170       180
orf7-1.pep    TPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRR
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf7ng-1      TPDIGHDTKGWSNEKLMAEVAPDAFSGNPEGQFFPDSYEIDAGGSDLQIYQTAYKAMQRR
                     100       110       120       130       140       150

                    190       200       210       220       230       240
orf7-1.pep    LNEAWESRQDGLPYKNPYEMLIMASLVEKETGHEADRDHVASVFVNRLKIGMRLQTDPSV
              ||||| :|||||||||||||||||||:|||||||||||||||||||||||||||||||||
orf7ng-1      LNEAWAGRQDGLPYKNPYEMLIMASLIEKETGHEADRDHVASVFVNRLKIGMRLQTDPSV
                     160       170       180       190       200       210

                    250       260       270       280       290       300
orf7-1.pep    IYGMGAAYKGKIRKADLRRDTPYNTYTRGGLPPTPIALPGKAALDAAAHPSGEKYLYFVS
```

```
              ||||||||||||||||||||||||| |||||| ||||||||:||||||||||||||
orf7ng-1      IYGMGAAYKGKIRKADLRRDTPYNTYTGGGLPPTRIALPGKAAMDAAAHPSGEKYLYFVS
                    220       230       240       250       260       270

                  310       320       330
orf7-1.pep    KMDGTGLSQFSHDLTEHNAAVRKYILKKX
              |||||||||||||||||||||||||||||
orf7ng-1      KMDGTGLSQFSHDLTEHNAAVRKYILKKX
                    280       290
```

In addition, ORF7ng-1 shows significant homology with a hypothetical *E.coli* protein:

```
sp|P28306|YCEG_ECOLI HYPOTHETICAL 38.2 KD PROTEIN IN PABC-HOLB INTERGENIC REGION
gi|1787339 (AE000210) o340; 100% identical to fragment YCEG_ECOLI SW: P28306 but
has 97 additional C-terminal residues [Escherichia coli] Length = 340
 Score = 79 (36.2 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57
 Identities = 20/87 (22%), Positives = 40/87 (45%)

Query:     10 GISSVGRKLAEDRIVFSRHVLTAAAYVLGVHNRLHTGTYRLPSEVSAWDILQKMRGGRPD 69
              G  ++G +L  D+I+     V      +     +     GTYR    +++    ++L+ +  G+
Sbjct:     49 GRLALGEQLYADKIINRPRVFQWLLRIEPDLSHFKAGTYRFTPQMTVREMLKLLESGKEA 108

Query:     70 SVTVQIIEGSRFSHMRKVIDATPDIGH 96
                 ++++EG R S   K +    P I H
Sbjct:    109 QFPLRLVEGMRLSDYLKQLREAPYIKH 135

 Score = 438 (200.7 bits), Expect = 5.0e-57, Sum P(2) = 5.0e-57
 Identities = 84/155 (54%), Positives = 111/155 (71%)

Query:    120 EGQFFPDSYEIDAGGSDLQIYQTAYKAMQRRLNEAWAGRQDGLPYKNPYEMLIMASLIEK 179
              EG F+PD++    A   +D+ + + A+K M + ++ AW GR DGLPYK+   +++ MAS+IEK
Sbjct:    158 EGWFWPDTWMYTANTTDVALLKRAHKKMVKAVDSAWEGRADGLPYKDKNQLVTMASIIEK 217

Query:    180 ETGHEADRDHVASVFVNRLKIGMRLQTDPSVIYGMGAAYKGKIRKADLRRDTPYNTYTGG 239
              ET    ++RD VASVF+NRL+IGMRLQTDP+VIYGMG   Y GK+ +ADL    T YNTYT
Sbjct:    218 ETAVASERDKVASVFINRLRIGMRLQTDPTVIYGMGERYNGKLSRADLETPTAYNTYTIT 277

Query:    240 GLPPTRIALPGKAAMDAAAHPSGEKYLYFVSKMDG 274
              GLPP  IA PG  ++ AAAHP+    YLYFV+   G
Sbjct:    278 GLPPGAIATPGADSLKAAAHPAKTPYLYFVADGKG 312
```

Based on this analysis, including the fact that the *H.influenzae* YCEG protein possesses a possible leader sequence, it is predicted that the proteins from *N meningitidis* and *N.gonorrhoeae,* and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 6

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 39>:

```
  1 CGTTTCAAAA TGTTAACTGT GTTGACGGCA ACCTTGATTG CCGGACAGGT
 51 ATCTGCCGCC GGAGGCGGTG CGGGGGATAT GAAACAGCCG AAGGAAGTCG
101 GAAAGGTTTT CAGAAAGCAG CAGCGTTACA GCGAGGAAGA AATCAAAAAC
151 GAACGCGCAC GGCTTGCGGC AGTGGGCGAG CGGGTTAATC AGATATTTAC
201 GTTGCTGGGA GGGGAAACCG CCTTGCAAAA GGGGCAGGCG GGAACGGCTC
251 TGGCAACCTA TATGCTGATG TTGGAACGCA CAAAATCCCC CGAAGTCGCC
301 GAACGCGCCT TGGAAATGGC CGTGTCGCTG AACGCGTTTG AACAGGCGGA
351 AATGATTTAT CAGAAATGGC GGCAGATTGA GCCTATACCG GGTAAGGCGC
401 AAAAACGGGC GGGGTGGCTG CGGAACGTGC TGAGGGAAAG AGGAAATCAG
451 CATCTGGACG GACGGGAAGA AGTGCTGGCT CAGGCGGACG AAGGACAG
```

This corresponds to the amino acid sequence <SEQ ID 40; ORF9>:

```
  1 ..RFKMLTVLTA TLIAGQVSAA GGGAGDMKQP KEVGKVFRKQ QRYSEEEIKN
 51   ERARLAAVGE RVNQIFTLLG GETALQKGQA GTALATYMLM LERTKSPEVA
101   ERALEMAVSL NAFEQAEMIY QKWRQIEPIP GKAQKRAGWL RNVLRERGNQ
```

```
151    HLDGREEVLA QADEGQ
```

Further sequence analysis revealed the complete DNA sequence <SEQ ID 41>:

```
   1  ATGTTACCTA ACCGTTTCAA AATGTTAACT GTGTTGACGG CAACCTTGAT
  51  TGCCGGACAG GTATCTGCCG CCGGAGGCGG TGCGGGGGAT ATGAAACAGC
 101  CGAAGGAAGT CGGAAAGGTT TTCAGAAAGC AGCAGCGTTA CAGCGAGGAA
 151  GAAATCAAAA ACGAACGCGC ACGGCTTGCG GCAGTGGGCG AGCGGGTTAA
 201  TCAGATATTT ACGTTGCTGG GAGGGGAAAC CGCCTTGCAA AAGGGGCAGG
 251  CGGGAACGGC TCTGGCAACC TATATGCTGA TGTTGGAACG CACAAAATCC
 301  CCCGAAGTCG CCGAACGCGC CTTGGAAATG GCCGTGTCGC TGAACGCGTT
 351  TGAACAGGCG GAAATGATTT ATCAGAAATG GCGGCAGATT GAGCCTATAC
 401  CGGGTAAGGC GCAAAAACGG GCGGGGTGGC TGCGGAACGT GCTGAGGGAA
 451  AGAGGAAATC AGCATCTGGA CGGACTGGAA GAAGTGCTGG CTCAGGCGGA
 501  CGAAGGACAG AACCGCAGGG TGTTTTTATT GTTGGCACAA GCCGCCGTGC
 551  AACAGGACGG GTTGGCGCAA AAAGCATCGA AAGCGGTTCG CCGCGCGGCG
 601  TTGAAATATG AACATCTGCC CGAAGCGGCG GTTGCCGATG TGGTGTTCAG
 651  CGTACAGGGA CGCGAAAAGG AAAAGGCAAT CGGAGCTTTG CAGCGTTTGG
 701  CGAAGCTCGA TACGGAAATA TTGCCCCCCA CTTTAATGAC GTTGCGTCTG
 751  ACTGCACGCA AATATCCCGA AATACTCGAC GGCTTTTTCG AGCAGACAGA
 801  CACCCAAAAC CTTTCGGCCG TCTGGCAGGA AATGGAAATT ATGAATCTGG
 851  TTTCCCTGCA CAGGCTGGAT GATGCCTATG CGCGTTTGAA CGTGCTGTTG
 901  GAACGCAATC CGAATGCAGA CCTGTATATT CAGGCAGCGA TATTGGCGGC
 951  AAACCGAAAA GAAGGTGCTT CCGTTATCGA CGGCTACGCC GAAAAGGCAT
1001  ACGGCAGGGG GACGGAGGAA CAGCGGAGCA GGGCGGCGCT AACGGCGGCG
1051  ATGATGTATG CCGACCGCAG GGATTACGCC AAAGTCAGGC AGTGGCTGAA
1101  AAAAGTATCC GCGCCGGAAT ACCTGTTCGA CAAAGGTGTG CTGGCGGCTG
1151  CGGCGGCTGT CGAGTTGGAC GGCGGCAGGG CGGCTTTGCG GCAGATCGGC
1201  AGGGTGCGGA AACTTCCCGA ACAGCAGGGG CGGTATTTTA CGGCAGACAA
1251  TTTGTCCAAA ATACAGATGC TCGCCCTGTC GAAGCTGCCC GATAAACGGG
1301  AGGCTTTGAG GGGGTTGGAC AAGATTATCG AAAAACCGCC TGCCGGCAGT
1351  AATACAGAGT TACAGGCAGA GGCATTGGTA CAGCGGTCAG TTGTTTACGA
1401  TCGGCTTGGC AAGCGGAAAA AAATGATTTC AGATCTTGAA AGGGCGTTCA
1451  GGCTTGCACC CGATAACGCT CAGATTATGA ATAATCTGGG CTACAGCCTG
1501  CTGACCGATT CCAAACGTTT GGACGAAGGT TTCGCCCTGC TTCAGACGGC
1551  ATACCAAATC AACCCGGACG ATACCGCTGT CAACGACAGC ATAGGCTGGG
1601  CGTATTACCT GAAAGGCGAC GCGGAAAGCG CGCTGCCGTA TCTGCGGTAT
1651  TCGTTTGAAA ACGACCCCGA GCCCGAAGTT GCCGCCCATT TGGGCGAAGT
1701  GTTGTGGGCA TTGGGCGAAC GCGATCAGGC GGTTGACGTA TGGACGCAGG
1751  CGGCACACCT TACGGGAGAC AAGAAAATAT GGCGGGAAAC GCTCAAACGT
1801  CACGGCATCG CATTGCCCCA ACCTTCCCGA AAACCTCGGA AATAA
```

This corresponds to the amino acid sequence <SEQ ID 42; ORF9-1>:

```
  1  MLPNRFKMLT VLTATLIAGQ VSAAGGGAGD MKQPKEVGKV FRKQQRYSEE
 51  EIKNERARLA AVGERVNQIF TLLGGETALQ KGQAGTALAT YMLMLERTKS
101  PEVAERALEM AVSLNAFEQA EMIYQKWRQI EPIPGKAQKR AGWLRNVLRE
151  RGNQHLDGLE EVLAQADEGQ NRRVFLLLAQ AAVQQDGLAQ KASKAVRRAA
201  LKYEHLPEAA VADVVFSVQG REKEKAIGAL QRLAKLDTEI LPPTLMTLRL
251  TARKYPEILD GFFEQTDTQN LSAVWQEMEI MNLVSLHRLD DAYARLNVLL
301  ERNPNADLYI QAAILAANRK EGASVIDGYA EKAYGRGTEE QRSRAALTAA
351  MMYADRRDYA KVRQWLKKVS APEYLFDKGV LAAAAAVELD GGRAALRQIG
401  RVRKLPEQQG RYFTADNLSK IQMLALSKLP DKREALRGLD KIIEKPPAGS
451  NTELQAEALV QRSVVYDRLG KRKKMISDLE RAFRLAPDNA QIMNNLGYSL
501  LTDSKRLDEG FALLQTAYQI NPDDTAVNDS IGWAYYLKGD AESALPYLRY
551  SFENDPEPEV AAHLGEVLWA LGERDQAVDV WTQAAHLTGD KKIWRETLKR
601  HGIALPQPSR KPRK*
```

Computer analysis of this amino acid sequence gave the following results:

<u>Homology with a predicted ORF from *N.meningitidis* (strain A)</u>

ORF9 shows 89.8% identity over a 166aa overlap with an ORF (ORF9a) from strain A of *N. meningitidis*:

```
                          10        20        30        40        50
orf9.pep        RFKMLTVLTATLIAGQVSAAGGGAGDMKQPKEVGKVFRKQQRYSEEEIKNERARLA
                || :|:||:|:|:|||: ||    ||:| | |||||||||||||||||||||||||
orf9a       MLPARFTILSVLAAALLAGQAYAA--GAADAKPPKEVGKVFRKQQRYSEEEIKNERARLA
```

```
                   10        20        30        40        50

                   60        70        80        90       100       110
orf9.pep   AVGERVNQIFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
           |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
orf9a      AVGERVNQIFTLLGXETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
           60        70        80        90       100       110

                  120       130       140       150       160
orf9.pep   EMIYQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGREEVLAQADEGQ
           |||||||||||||||||||||||||||||||||||||| || |||||| |
orf9a      EMIYQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGLEEXLAQADEXQNRRVFLLLAQ
          120       130       140       150       160       170

orf9a      AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVVFSVQXREKEKAIGALQRLAKLDTEI
          180       190       200       210       220       230
```

The complete length ORF9a nucleotide sequence <SEQ ID 43> is:

```
   1  ATGTTACCCG CCCGTTTCAC CATTTTATCT GTGCTCGCGG CAGCCCTGCT
  51  TGCCGGGCAG GCGTATGCCG CCGGCGCGGC GGATGCGAAG CCGCCGAAGG
 101  AAGTCGGAAA GGTTTTCAGA AAGCAGCAGC GTTACAGCGA GGAAGAAATC
 151  AAAAACGAAC GCGCACGGCT TGCGGCAGTG GGCGAGCGGG TTAATCAGAT
 201  ATTTACGTTG CTGGGANGGG AAACCGCCTT GCAAAAGGGG CAGGCGGGAA
 251  CGGCTCTGGC AACCTATATG CTGATGTTGG AACGCACAAA ATCCCCCGAA
 301  GTCGCCGAAC GCGCCTTGGA AATGGCCGTG TCNCTGAACG CGTTTGAACA
 351  GGCGGAAATG ATTTATCAGA AATGGCGGCA GATTGAGCCT ATACCGGGTA
 401  AGGCGCAAAA ACGGGCGGGG TGGCTGCGGA ACGTGCTGAG GGAAAGAGGA
 451  AATCAGCATC TAGACGGACT GGAAGAANTG CTGGCTCAGG CGGACGAANG
 501  ACAGAACCGC AGGGTGTTTT TATTGTTGGC ACAAGCCGCC GTGCAACAGG
 551  ACGGGTTGGC GCAAAAAGCA TCGAAAGCGG TTCGCCGCGC GGCGTTGAGA
 601  TATGAACATC TGCCCGAAGC GGCGGTTGCC GATGTGGTGT TCAGCGTACA
 651  GGNACGCGAA AAGGAAAAGG CAATCGGAGC TTTGCAGCGT TTGGCGAAGC
 701  TCGATACGGA AATATTGCCC CCCACTTTAA TGACGTTGCG TCTGACTGCA
 751  CGCAAATATC CCGAAATACT CGACGGCTTT TTCGAGCAGA CAGACACCCA
 801  AAACCTTTCG GCCGTCTGGC AGGAAATGGA AATTATGAAT CTGGTTTCCC
 851  TGCACAGGCT GGATGATGCC TATGCGCGTT TGAACGTGCT GTTGGAACGC
 901  AATCCGAATG CAGACCTGTA TATTCAGGCA GCGATATTGG CGGCAAACCG
 951  AAAAGAANGT GCTTCCGTTA TCGACGGCTA CGCCGAAAAG GCATACGGCA
1001  GGGGGACGGG GGAACAGCGG GGCAGGGCGG CAATGACGGC GGCGATGATA
1051  TATGCCGACC GAAGGGATTA CACCAAAGTC AGGCAGTGGT TGAAAAAAGT
1101  GTCCGCGCCG GAATACCTGT TCGACAAAGG TGTGCTGGCG GCTGCGGCGG
1151  CTGTCGAGTT GGACNGCGGC AGGGCGGCTT TGCGGCAGAT CGGCAGGGTG
1201  CGGAAACTTC CCGAACAGCA GGGGCGGTAT TTTACGGCAG ACAATTTGTC
1251  CAAAATACAG ATGTTCGCCC TGTCGAAGCT GCCCGACAAA CGGGAGGCTT
1301  TGAGGGGGTT GGACAAGATT ATCGAAAAAC CGCCTGCCGG CAGTAATACA
1351  GAGTTACAGG CAGAGGCATT GGTACAGCGG TCAGTTGTTT ACGATCGGCT
1401  TGGCAAGCGG AAAAAAATGA TTTCAGATCT TGAAAGGGCG TTCAGGCTTG
1451  CACCCGATAA CGCTCAGATT ATGAATAATC TGGGCTACAG CCTGCTTTCC
1501  GATTCCAAAC GTTTGGACGA AGGCTTCGCC CTGCTTCAGA CGGCATACCA
1551  AATCAACCCG GACGATACCG CTGTCAACGA CAGCATAGGC TGGGCGTATT
1601  ACCTGAAANG CGACGCGGAA AGCGCGCTGC CGTATCTGCG GTATTCGTTT
1651  GAAAACGACC CCGAGCCCGA AGTTGCCGCC CATTTGGGCG AAGTGTTGTG
1701  GGCATTGGGC GAACGCGATC AGGCGGTTGA CGTATGGACG CAGGCGGCAC
1751  ACCTTACGGG AGACAAGAAA ATATGGCGGG AAACGCTCAA ACGTCACGGC
1801  ATCGCATTGC CCCAACCTTC CCGAAAACCT CGGAAATAA
```

This encodes a protein having amino acid sequence <SEQ ID 44>:

```
  1  MLPARFTILS VLAAALLAGQ AYAAGAADAK PPKEVGKVFR KQQRYSEEEI
 51  KNERARLAAV GERVNQIFTL LGXETALQKG QAGTALATYM LMLERTKSPE
101  VAERALEMAV SLNAFEQAEM IYQKWRQIEP IPGKAQKRAG WLRNVLRERG
151  NQHLDGLEEX LAQADEXQNR RVFLLLAQAA VQQDGLAQKA SKAVRRAALR
201  YEHLPEAAVA DVVFSVQXRE KEKAIGALQR LAKLDTEILP PTLMTLRLTA
251  RKYPEILDGF FEQTDTQNLS AVWQEMEIMN LVSLHRLDDA YARLNVLLER
301  NPNADLYIQA AILAANRKEX ASVIDGYAEK AYGRGTGEQR GRAAMTAAMI
351  YADRRDYTKV RQWLKKVSAP EYLFDKGVLA AAAAVELDXG RAALRQIGRV
401  RKLPEQQGRY FTADNLSKIQ MFALSKLPDK REALRGLDKI IEKPPAGSNT
451  ELQAEALVQR SVVYDRLGKR KKMISDLERA FRLAPDNAQI MNNLGYSLLS
501  DSKRLDEGFA LLQTAYQINP DDTAVNDSIG WAYYLKXDAE SALPYLRYSF
551  ENDPEPEVAA HLGEVLWALG ERDQAVDVWT QAAHLTGDKK IWRETLKRHG
```

```
601  IALPQPSRKP RK*
```

ORF9a and ORF9-1 show 95.3% identity in 614 aa overlap:

```
                    10        20        30        40        50
orf9a.pep   MLPARFTILSVLAAALLAGQAYAAG--AADAKPPKEVGKVFRKQQRYSEEEIKNERARLA
            ||| || :|:||:|:|:|||: |||  |:| | ||||||||||||||||||||||||||
orf9-1      MLPNRFKMLTVLTATLIAGQVSAAGGGAGDMKQPKEVGKVFRKQQRYSEEEIKNERARLA
                    10        20        30        40        50        60

            60        70        80        90       100       110
orf9a.pep   AVGERVNQIFTLLGXETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
            ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
orf9-1      AVGERVNQIFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
                    70        80        90       100       110       120

           120       130       140       150       160       170
orf9a.pep   EMIYQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGLEEXLAQADEXQNRRVFLLLAQ
            ||||||||||||||||||||||||||||||||||||||||| |||||| ||||||||||
orf9-1      EMIYQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGLEEVLAQADEGQNRRVFLLLAQ
                   130       140       150       160       170       180

           180       190       200       210       220       230
orf9a.pep   AAVQQDGLAQKASKAVRRAALRYEHLPEAAVADVVFSVQXREKEKAIGALQRLAKLDTEI
            ||||||||||||||||||||:|||||||||||||||||| ||||||||||||||||||||
orf9-1      AAVQQDGLAQKASKAVRRAALKYEHLPEAAVADVVFSVQGREKEKAIGALQRLAKLDTEI
                   190       200       210       220       230       240

           240       250       260       270       280       290
orf9a.pep   LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLHRLDDAYARLNVLL
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf9-1      LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLHRLDDAYARLNVLL
                   250       260       270       280       290       300

           300       310       320       330       340       350
orf9a.pep   ERNPNADLYIQAAILAANRKEXASVIDGYAEKAYGRGTGEQRGRAAMTAAMIYADRRDYT
            |||||||||||||||||||| |||||||||||||||| |||:|||:||||:|||||||:
orf9-1      ERNPNADLYIQAAILAANRKEGASVIDGYAEKAYGRGTEEQRSRAALTAAMMYADRRDYA
                   310       320       330       340       350       360

           360       370       380       390       400       410
orf9a.pep   KVRQWLKKVSAPEYLFDKGVLAAAAAVELDXGRAALRQIGRVRKLPEQQGRYFTADNLSK
            |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
orf9-1      KVRQWLKKVSAPEYLFDKGVLAAAAAVELDGGRAALRQIGRVRKLPEQQGRYFTADNLSK
                   370       380       390       400       410       420

           420       430       440       450       460       470
orf9a.pep   IQMFALSKLPDKREALRGLDKIIEKPPAGSNTELQAEALVQRSVVYDRLGKRKKMISDLE
            |||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf9-1      IQMLALSKLPDKREALRGLDKIIEKPPAGSNTELQAEALVQRSVVYDRLGKRKKMISDLE
                   430       440       450       460       470       480

           480       490       500       510       520       530
orf9a.pep   RAFRLAPDNAQIMNNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKXD
            |||||||||||||||||||||:|||||||||||||||||||||||||||||||||||| |
orf9-1      RAFRLAPDNAQIMNNLGYSLLTDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD
                   490       500       510       520       530       540

           540       550       560       570       580       590
orf9a.pep   AESALPYLRYSFENDPEPEVAAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLKR
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf9-1      AESALPYLRYSFENDPEPEVAAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLKR
                   550       560       570       580       590       600

           600       610
orf9a.pep   HGIALPQPSRKPRKX
            ||||||||||||||
orf9-1      HGIALPQPSRKPRKX
                   610
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF9 shows 82.8% identity over a 163aa overlap with a predicted ORF (ORF9.ng) from *N. gonorrhoeae:*

```
Orf9             RFKMLTVLTATLIAGQVSAAGGGAGDMKQPKEVGKVFRKQQRYSEEEIKNERAR   54
                 || :|:||:|:|:|||: ||   ||:|:: |||||||:||::||||||||||||
orf9ng       MIMLPARFTILSVLAAALLAGQAYAA--GAADVELPKEVGKVLRKHRRYSEEEIKNERAR   58

orf9         LAAVGERVNQIFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE  114
             |||||||||::|||||||||||||||||||||||||||||||||||||||||||||||||
orf9ng       LAAVGERVNRVFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFE  118

orf9         QAEMIYQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGREEVLAQADEGQ         166
             |||||||||||||||||:||| ||||||||:| || |||   ||| ||:|
orf9ng       QAEMIYQKWRQIEPIPGEAQKPAGWLRNVLKEGGNPHLDRLEEVPAQSDYVHQPMIFLLL  178
```

The ORF9ng nucleotide sequence <SEQ ID 45> was predicted to encode a protein having including acid sequence <SEQ ID 46>:

```
  1 MIMLPARFTI LSVLAAALLA GQAYAAGAAD VELPKEVGKV LRKHRRYSEE
 51 EIKNERARLA AVGERVNRVF TLLGGETALQ KGQAGTALAT YMLMLERTKS
101 PEVAERALEM AVSLNAFEQA EMIYQKWRQI EPIPGEAQKP AGWLRNVLKE
151 GGNPHLDRLE EVPAQSDYVH QPMIFLLLVQ AAVQHGGVAQ KPSKAVRPAA
201 YNYEVLPETA GADAVFCVQG PQYEKAIQSF PPCGRNPQTE NIAPPFNELF
251 RPTARPISPK LLQRFFRTEP NLAKPFRPPG PEMETYQTGF PRPLTRNNPT
```

Amino acids 1-28 are a putative leader sequence, and 173-189 are predicted to be a transmembrane domain.

Further sequence analysis revealed the complete length ORF9ng DNA sequence <SEQ ID 47>:

```
   1 ATGTTACCCG CCCGTTTCAC TATTTTATCT GTCCTCGCAG CAGCCCTGCT
  51 TGCCGGACAG GCGTATGCTG CCGGCGCGGC GGATGTGGAG CTGCCGAAGG
 101 AAGTCGGAAA GGTTTTAAGG AAACATCGGC GTTACAGCGA GGAAGAAATC
 151 AAAAACGAAC GCGCACGGCT TGCGGCAGTG GGCGAACGGG TCAACAGGGT
 201 GTTTACGCTG TTGGGCGGTG AAACGGCTTT GCAGAAAGGG CAGGCGGGAA
 251 CGGCTCTGGC AACCTATATG CTGATGTTGG AACGCACAAA ATCCCCCGAA
 301 GTCGCCGAAC GCGCCTTGGA AATGGCCGTG TCGCTGAACG CGTTTGAACA
 351 GGCGGAAATG ATTTATCAGA AATGgcggca gatcgagcct ataCcgggtg
 401 aggcgcaaaa accgGcgggG tggctgcgga acgtattgaa ggaagggGGa
 451 aaTCAGCATC TGGAcgggtt gaaagaggTG CtggcgcaAT cggacgatGT
 501 GCAAAAAcgc aggaTATTTT TGCTGCTGGT GCAAGCCGCC GTGCagcagg
 551 gTGGGGTGGC TCAAAAAGCA TCGAAAGCGG TTCGCcgtgc GGcgttgaAG
 601 TATGAACATC TGCCcgaagc ggcggTTGCC GATGcggTGT TCGGCGTACA
 651 GGGACGCGAA AAGGAAAagg caaTCGAAGC TTTGCAGCGT TTGGCGAAGC
 701 TCGATACGGA AATATTGCCC CCCACTTTAA TGACGTTGCG TCTGACTGCA
 751 CGCAAATATC CCGAAATACT CGACGGCTTT TTCGAGCAGA CAGACACCCA
 801 AAACCTTTCG GCCGTCTGGC AGGAAATGGA AATTATGAAT CTGGTTTCCC
 851 TGCGTAAGCC GGATGATGCC TATGCGCGTT TGAACGTGCT GTTGGAACAC
 901 AACCCGAATG CAAACCTGTA TATTCAGGCG GCGATATTGG CGGCAAACCG
 951 AAAAGAAGGT GCGTCCGTTA TCGACGGCTA CGCCGAAAAG GCATACGGCA
1001 GGGGGACGGG GGAACAGCGG GGCagggcgg cAATgacggc GGCGATGATA
1051 TATGCCGACC GCAGGGATTA CGCCAAAGTC AGGCAGTGGT TGAAAAAAGT
1101 GTCCGCGCCG GAATACCTGT TCGACAAAGG CGTGCTGGCG GCTGCGGCGG
1151 CTGCCGAATT GGACGGAGGC CGGGCGGCTT TGCGGCAGAT CGGCAGGGTG
1201 CGGAAACTTC CCGAACAGCA GGGGCGGTAT TTTACGGCAG ACAATTTGTC
1251 CAAAATACAG ATGCTCGCCC TGTCGAAGCT GCCCGACAAA CGGGAAGCCC
1301 TGATCGGGCT GAACAACATC ATCGCCAAAC TTTCGGCGGC GGGAAGCACG
1351 GAACCTTTGG CGGAAGCATT GGCACAGCGT TCCATTATTT ACGaacAGTT
1401 cggCAAACGG GGAAAAATGA TTGCCGACCT tgaAACcgcg CTCAAACTTA
1451 CGCCCGATAA TGCACAAATT ATGAATAATC TGGGCTACAG CCTGCTTTCC
1501 GATTCCAAAC GTTTGGACGA GGGTTTCGCC CTGCTTCAGA CGGCATACCA
1551 AATCAACCCG GACGATACCG CCGTTAACGA CAGCATAGGC TGGGCGTATT
1601 ACCTGAAAGG CGACgcggaA AGCGCGCTGC CGTATCTGcg gtattcgttt
1651 gAAAACGACC CCGAGCCCGA AGTTGCCGCC CATTTGGGCG AAGTGTTGTG
```

```
1701  GGCATTGGGC GAACGCGATC AGGCGGTTGA CGTATGGACG CAGGCGGCAC
1751  ACCTTAGGGG AGACAAGAAA ATATGGCGGG AGACGCTCAA ACGCTACGGA
1801  ATCGCCTTGC CCGAGCCTTC CCGAAAACCC CGGAAATAA
```

This encodes a protein having amino acid sequence <SEQ ID 48>:

```
  1  MLPARFTILS VLAAALLAGQ AYAAGAADVE LPKEVGKVLR KHRRYSEEEI
 51  KNERARLAAV GERVNRVFTL LGGETALQKG QAGTALATYM LMLERTKSPE
101  VAERALEMAV SLNAFEQAEM IYQKWRQIEP IPGEAQKPAG WLRNVLKEGG
151  NQHLDGLKEV LAQSDDVQKR RIFLLLVQAA VQQGGVAQKA SKAVRRAALK
201  YEHLPEAAVA DAVFGVQGRE KEKAIEALQR LAKLDTEILP PTLMTLRLTA
251  RKYPEILDGF FEQTDTQNLS AVWQEMEIMN LVSLRKPDDA YARLNVLLEH
301  NPNANLYIQA AILAANRKEG ASVIDGYAEK AYGRGTGEQR GRAAMTAAMI
351  YADRRDYAKV RQWLKKVSAP EYLFDKGVLA AAAAAELDGG RAALRQIGRV
401  RKLPEQQGRY FTADNLSKIQ MLALSKLPDK REALIGLNNI IAKLSAAGST
451  EPLAEALAQR SIIYEQFGKR GKMIADLETA LKLTPDNAQI MNNLGYSLLS
501  DSKRLDEGFA LLQTAYQINP DDTAVNDSIG WAYYLKGDAE SALPYLRYSF
551  ENDPEPEVAA HLGEVLWALG ERDQAVDVWT QAAHLRGDKK IWRETLKRYG
601  IALPEPSRKP RK*
```

ORF9ng and ORF9-1 show 88.1% identity in 614 aa overlap:

```
                    10        20        30        40        50        60
orf9-1.pep  MLPNRFKMLTVLTATLIAGQVSAAGGGAGDMKQPKEVGKVFRKQQRYSEEEIKNERARLA
            ||| || :|:||:|:|:|||: |||    |:|:: ||||||:||::||||||||||||||
orf9ng-1    MLPARFTILSVLAAALLAGQAYAAG--AADVELPKEVGKVLRKHRRYSEEEIKNERARLA
                    10        20          30        40        50

                    70        80        90       100       110       120
orf9-1.pep  AVGERVNQIFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
            ||||||::||||||||||||||||||||||||||||||||||||||||||||||||||||
orf9ng-1    AVGERVNRVFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQA
          60        70        80        90       100       110

                   130       140       150       160       170       180
orf9-1.pep  EMIYQKWRQIEPIPGKAQKRAGWLRNVLRERGNQHLDGLEEVLAQADEGQNRRVFLLLAQ
            |||||||||||||||:||| ||||||||:| |||||||:|||||:|: |:||:||||:|
orf9ng-1    EMIYQKWRQIEPIPGEAQKPAGWLRNVLKEGGNQHLDGLKEVLAQSDDVQKRRIFLLLVQ
         120       130       140       150       160       170

                   190       200       210       220       230       240
orf9-1.pep  AAVQQDGLAQKASKAVRRAALKYEHLPEAAVADVVFSVQGREKEKAIGALQRLAKLDTEI
            ||||| |:||||||||||||||||||||||||||:||:|||||||||| |||||||||||
orf9ng-1    AAVQQGGVAQKASKAVRRAALKYEHLPEAAVADAVFGVQGREKEKAIEALQRLAKLDTEI
         180       190       200       210       220       230

                   250       260       270       280       290       300
orf9-1.pep  LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLHRLDDAYARLNVLL
            |||||||||||||||||||||||||||||||||||||||||||||||:: ||||||||||
orf9ng-1    LPPTLMTLRLTARKYPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLRKPDDAYARLNVLL
         240       250       260       270       280       290

                   310       320       330       340       350       360
orf9-1.pep  ERNPNADLYIQAAILAANRKEGASVIDGYAEKAYGRGTEEQRSRAALTAAMMYADRRDYA
            |:||||:|||||||||||||||||||||||||||||||| |||:|||:||||:|||||||
orf9ng-1    EHNPNANLYIQAAILAANRKEGASVIDGYAEKAYGRGTGEQRGRAAMTAAMIYADRRDYA
         300       310       320       330       340       350

                   370       380       390       400       410       420
orf9-1.pep  KVRQWLKKVSAPEYLFDKGVLAAAAAVELDGGRAALRQIGRVRKLPEQQGRYFTADNLSK
            ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
orf9ng-1    KVRQWLKKVSAPEYLFDKGVLAAAAAAELDGGRAALRQIGRVRKLPEQQGRYFTADNLSK
         360       370       380       390       400       410

                   430       440       450       460       470       480
orf9-1.pep  IQMLALSKLPDKREALRGLDKIIEKPPAGSNTELQAEALVQRSVVYDRLGKRKKMISDLE
            ||||||||||||||| ||::|| |   |:::||  ||||:|||::|:::||| |||:|||
orf9ng-1    IQMLALSKLPDKREALIGLNNIIAKLSAAGSTEPLAEALAQRSIIYEQFGKRGKMIADLE
         420       430       440       450       460       470

                   490       500       510       520       530       540
```

```
                   550       560       570       580       590       600
orf9-1.pep   RAFRLAPDNAQIMNNLGYSLLTDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD
              |::|:|||||||||||||||:||||||||||||||||||||||||||||||||||||||
orf9ng-1     TALKLTPDNAQIMNNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGD
            480       490       500       510       520       530

                     550       560       570       580       590       600
orf9-1.pep   AESALPYLRYSFENDPEPEVAAHLGEVLWALGERDQAVDVWTQAAHLTGDKKIWRETLKR
             ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
orf9ng-1     AESALPYLRYSFENDPEPEVAAHLGEVLWALGERDQAVDVWTQAAHLRGDKKIWRETLKR
            540       550       560       570       580       590

                     610
orf9-1.pep   HGIALPQPSRKPRKX
             :|||||:||||||||
orf9ng-1     YGIALPEPSRKPRKX
            600       610
```

In addition, ORF9ng shows significant homology with a hypothetical protein from *P.aeruginosa*:

```
sp|P42810|YHE3_PSEAE HYPOTHETICAL 64.8 KD PROTEIN IN HEMM-HEMA INTERGENIC REGION
(ORF3)
>gi|1072999|pir||S49376 hypothetical protein 3 - Pseudomonas aeruginosa >gi|557259
(X82071) orf3  [Pseudomonas aeruginosa] Length = 576
 Score =  128 bits (318), Expect = 1e-28
 Identities = 138/587 (23%), Positives = 228/587 (38%), Gaps = 125/587 (21%)

Query: 67  VFTLLGGETALQKGQAGTALATYMLMLERTKSPEVAERALEMAVSLNAFEQAEMIYQKWR 126
            +++LL  E A Q+ +    AL+ Y++   ++T+ P V+ERA  +A  L A ++A      W
Sbjct: 53  LYSLLVAELAGQRNRFDIALSNYVVQAQKTRDPGVSERAFRIAEYLGADQEALDTSLLWA 112

Query: 127 QIEPIPGEAQKPAG--------------WLRNVLKEGGNQHLDGLKEVLAQSDDVQKRRI 172
            +  P   +AQ+ A               ++  VL   G+ H D L    A++D   +  +
Sbjct: 113 RSAPDNLDAQRAAAIQLARAGRYEESMVYMEKVLNGQGDTHFDFLALSAAETDPDTRAGL 172

Query: 173 FXXXXXXXXXXXXXXXKASKAVRRAALKYEHLPEAAVADAVFGVQGREKEKAIEALQRLA 232
                            ++       KY +  +     A+   Q     ++A+   L+  +
Sbjct: 173 L-----------------QSFDHLLKKYPNNGQLLFGKALLLQQDGRPDEALTLLEDNS 214

Query: 233 KLDTEILPPTLMTLRLTARK-----YPEILDGFFEQTDTQNLSAVWQEMEIMNLVSLRKP 287
               E+ P  L +  L + K       P +  G  E  D + +    +  +     LV   +
Sbjct: 215 ASRHEVAPLLLRSRLLQSMKRSDEALPLLKAGIKEHPDDKRVRLAYARL----LVEQNRL 270

Query: 288 DDAYARLNVLLEHNPN---------------------ANLYIQAAI-------------- 312
            DDA A     L++  P+                      A +Y++  +
Sbjct: 271 DDAKAEFAGLVQQFPDDDDDLRFSLALVCLEAQAWDEARIYLEELVERDSHVDAAHFNLG 330

Query: 313 -LAANRKEGASVIDGYAEKAYGRGTGEQRGRAAMTAAMIYADRRDYAKVRQWLKKVSAPE 371
             LA   +K+ A   +D YA+    G G        +    T   ++ A R D A  R     +   P+
Sbjct: 331 RLAEEQKDTARALDEYAQ--VGPGNDFLPAQLRQTDVLLKAGRVDEAAQRLDKARSEQPD 388

Query: 372 YLFDKXXXXXXXXXXXXXXXXXXXXRQIGRVRKLPEQQGRYFTADNLSKIQMLALSKLPDKR 431
            Y                                        A  L  I+  ALS     +
Sbjct: 389 Y----------------------------------------AIQLYLIEAEALSNNDQQE 408

Query: 432 EALIGLNNIIAKLSAAGSTEPLAEALAQRSIIYEQFGKRGKMIADLETALKLTPDNAQIM 491
            +A    +    + +         E L    L  RS++ E+      +M   DL     +    PDNA   +
Sbjct: 409 KAWQAIQEGLKQYP-----EDL-NLLYTRSMLAEKRNDLAQMEKDLRFVIAREPDNAMAL 462

Query: 492 NNLGYSLLSDSKRLDEGFALLQTAYQINPDDTAVNDSIGWAYYLKGDAESALPYLRYSFE 551
            N LGY+L     + R   E    L+  A+++NPDD A+ DS+GW   Y +G    A   YLR + +
Sbjct: 463 NALGYTLADRTTRYGEARELILKAHKLNPDDPAILDSMGWINYRQGKLADAERYLRQALQ 522

Query: 552 NDPEPEVAAHLGEVLWALGERDQAVDVWTQAAHLRGDKKIWRETLKR 598
              P+ EVAAHLGEVLWA G +   A   +W +     + D   + R T+KR
Sbjct: 523 RYPDHEVAAHLGEVLWAQGRQGDARAIWREYLDKQPDSDVLRRTIKR 569

 gi|2983399 (AE000710) hypothetical protein [Aquifex aeolicus] Length = 545
 Score = 81.5 bits (198), Expect = 1e-14
 Identities = 61/198 (30%), Positives = 98/198 (48%), Gaps = 19/198 (9%)

Query: 408 GRYFTADNL-SKIQMLALSKLPDKREALIGLNNIIAKLSAAGSTEPLAEALAQ------- 459
            G Y  A  L  K ++LA    PDK+E L     +   +K         + + L +
```

```
Sbjct: 335 GNYEDAKRLIEKAKVLA----PDKKEILFLEADYYSKTKQYDKALEILKKLEKDYPNDSR 390

Query: 460 ----RSIIYEQFGKRGKMIADLETALKLTPDNAQIMNNLGYSLLS--DSKRLDEGFALLQ 513
               +I+Y+  G         L  A++L P+N    N LGYSLL     +R++E   L++
Sbjct: 391 VYFMEAIVYDNLGDIKNAEKALRKAIELDPENPDYYNYLGYSLLLWYGKERVEEAEELIK 450

Query: 514 TAYQINPDDTAVNDSIGWAYYLKGDAESALPYLRYSF-ENDPEPEVAAHLGEVLWALGER 572
              A + +P++ A  DS+GW YYLKGD E A+ YL  +  E   +P V  H+G+VL  +G +
Sbjct: 451 KALEKDPENPAYIDSMGWVYYLKGDYERAMQYLLKALREAYDDPVVNEHVGDVLLKMGYK 510

Query: 573 DQAVDVWTQAAHLRGDKK 590
           ++A + + +A  L  + K
Sbjct: 511 EEARNYYERALKLLEEGK 528
```

Based on this analysis, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 7

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 49>:

```
  1   AACCTCTACG CCGGCCCGCA GACCACATCC GTCATCGCAA ACATCGCCGA
 51   CAACCTGCAA CTGGCCAAAG ACTACGGCAA AGTACACTGG TTCGCCTCCC
101   CGCTCTTCTG GCTCCTGAAC CAACTGCACA ACATCATCGG CAACTGGGGC
151   TGGGCGATTA TCGTTTTAAC CATCATCGTC AAAGCCGTAC TGTATCCATT
201   GACCAACGCC TCTTACCGCT CTATGGCGAA AATGCGTGCC GCCGCACCCA
251   AACTGCAAGC CATCAAAGAG AAATACGGCG ACGACCGTAT GGCGCAACAA
301   CAGGCGATGA TGCAGCTTTA CACAGACGAG AAAATCAACC CGaCTGGGCG
351   GCTGCCTGCC TATGCTGTTG CAAATCCCCG TCTTCATCGG ATTGTATTGG
401   GCATTGTTCG CCTCCGTAGA ATTGCGCCAG GCACCTTGGC TGGGTTGGAT
451   TACCGACCTC AGCCGCGCCG ACCCCTACTA CATCCTGCCC ATCATTATGG
501   CGGCAACGAT GTTCGCCCAA ACTTATCTGA ACCCGCCGCC GAcCGACCCG
551   ATGCagGCGA AAATGATGAA AATCATGCCG TTGGTTTTCT CsGwCrTGTT
601   CTTCTTCTTC CCTGCCGGks TGGTATTGTA CTGGGTAGTC AACAACCTCC
651   TGACCATCGC CCAGCAATGG CACATCAACC GCAGCATCGA AAAACAACGC
701   GCCCAAGGCG AAGTCGTTTC CTAA
```

This corresponds to the amino acid sequence <SEQ ID 50; ORF11>:

```
  1   ..NLYAGPQTTS VIANIADNLQ LAKDYGKVHW FASPLFWLLN QLHNIIGNWG
 51     WAIIVLTIIV KAVLYPLTNA SYRSMAKMRA AAPKLQAIKE KYGDDRMAQQ
101     QAMMQLYTDE KINPLGGCLP MLLQIPVFIG LYWALFASVE LRQAPWLGWI
151     TDLSRADPYY ILPIIMAATM FAQTYLNPPP TDPMQAKMMK IMPLVFSXXF
201     FFFPAGXVLY WVVNNLLTIA QQWHINRSIE KQRAQGEVVS *
```

Further sequence analysis revealed the complete DNA sequence <SEQ ID 51>:

```
  1   ATGGATTTTA AAAGACTCAC GGCGTTTTTC GCCATCGCGC TGGTGATTAT
 51   GATCGGCTGG GAAAAGATGT TCCCCACTCC GAAGCCAGTC CCCGCGCCCC
101   AACAGGCAGC ACAACAACAG GCCGTAACCG CTTCCGCCGA AGCCGCGCTC
151   GCGCCCGCAA CGCCGATTAC CGTAACGACC GACACGGTTC AAGCCGTCAT
201   TGATGAAAAA AGCGGCGACC TGCGCCGGCT GACCCTGCTC AAATACAAAG
251   CAACCGGCGA CGAAAATAAA CCGTTCATCC TGTTTGGCGA CGGCAAAGAA
301   TACACCTACG TCGCCCAATC CGAACTTTTG GACGCGCAGG GCAACAACAT
351   TCTAAAAGGC ATCGGCTTTA GCGCACCGAA AAAACAGTAC AGCTTGGAAG
401   GCGACAAAGT TGAAGTCCGC CTGAGCGCGC CTGAAACACG CGGTCTGAAA
451   ATCGACAAAG TTTATACTTT CACCAAAGGC AGCTATCTGG TCAACGTCCG
501   CTTCGACATC GCCAACGGCA GCGGTCAAAC CGCCAACCTG AGCGCGGACT
551   ACCGCATCGT CCGCGACCAC AGCGAACCCG AGGGTCAAGG TTACTTTACC
601   CACTCTTACG TCGGCCCTGT TGTTTATACC CCTGAAGGCA ACTTCCAAAA
651   AGTCAGCTTT TCCGACTTGG ACGACGATGC CAAATCCGGC AAATCCGAGG
701   CCGAATACAT CCGCAAAACC CCGACCGGCT GGCTCGGCAT GATTGAACAC
751   CACTTCATGT CCACCTGGAT TCTCCAACCT AAAGGCAGAC AAAGCGTTTG
801   CGCCGCAGGC GAGTGCAACA TCGACATCAA ACGCCGCAAC GACAAGCTGT
851   ACAGCACCAG CGTCAGCGTG CCTTTAGCCG CCATCCAAAA CGGCGCGAAA
901   GCCGAAGCCT CCATCAACCT CTACGCCGGC CCGCAGACCA CATCCGTCAT
951   CGCAAACATC GCCGACAACC TGCAACTGGC CAAAGACTAC GGCAAAGTAC
```

```
1001  ACTGGTTCGC CTCCCCGCTC TTCTGGCTCC TGAACCAACT GCACAACATC
1051  ATCGGCAACT GGGGCTGGGC GATTATCGTT TTAACCATCA TCGTCAAAGC
1101  CGTACTGTAT CCATTGACCA ACGCCTCTTA CCGCTCTATG GCGAAAATGC
1151  GTGCCGCCGC ACCCAAACTG CAAGCCATCA AAGAGAAATA CGGCGACGAC
1201  CGTATGGCGC AACAACAGGC GATGATGCAG CTTTACACAG ACGAGAAAAT
1251  CAACCCGCTG GGCGGCTGCC TGCCTATGCT GTTGCAAATC CCCGTCTTCA
1301  TCGGATTGTA TTGGGCATTG TTCGCCTCCG TAGAATTGCG CCAGGCACCT
1351  TGGCTGGGTT GGATTACCGA CCTCAGCCGC GCCGACCCCT ACTACATCCT
1401  GCCCATCATT ATGGCGGCAA CGATGTTCGC CCAAACTTAT CTGAACCCGC
1451  CGCCGACCGA CCCGATGCAG GCGAAAATGA TGAAAATCAT GCCGTTGGTT
1501  TTCTCCGTCA TGTTCTTCTT CTTCCCTGCC GGTCTGGTAT TGTACTGGGT
1551  AGTCAACAAC CTCCTGACCA TCGCCCAGCA ATGGCACATC AACCGCAGCA
1601  TCGAAAAACA ACGCGCCCAA GGCGAAGTCG TTTCCTAA
```

This corresponds to the amino acid sequence <SEQ ID 52; ORF11-1>:

```
  1  MDFKRLTAFF AIALVIMIGW EKMFPTPKPV PAPQQAAQQQ AVTASAEAAL
 51  APATPITVTT DTVQAVIDEK SGDLRRLTLL KYKATGDENK PFILFGDGKE
101  YTYVAQSELL DAQGNNILKG IGFSAPKKQY SLEGDKVEVR LSAPETRGLK
151  IDKVYTFTKG SYLVNVRFDI ANGSGQTANL SADYRIVRDH SEPEGQGYFT
201  HSYVGPVVYT PEGNFQKVSF SDLDDDAKSG KSEAEYIRKT PTGWLGMIEH
251  HFMSTWILQP KGRQSVCAAG ECNIDIKRRN DKLYSTSVSV PLAAIQNGAK
301  AEASINLYAG PQTTSVIANI ADNLQLAKDY GKVHWFASPL FWLLNQLHNI
351  IGNWGWAIIV LTIIVKAVLY PLTNASYRSM AKMRAAAPKL QAIKEKYGDD
401  RMAQQQAMMQ LYTDEKINPL GGCLPMLLQI PVFIGLYWAL FASVELRQAP
451  WLGWITDLSR ADPYYILPII MAATMFAQTY LNPPPTDPMQ AKMMKIMPLV
501  FSVMFFFFPA GLVLYWVVNN LLTIAQQWHI NRSIEKQRAQ GEVVS*
```

Computer analysis of this amino acid sequence gave the following results:

<u>Homology with a 60kDa inner-membrane protein (accession P25754) of *Pseudomonas putida*</u>

ORF11 and the 60kDa protein show 58% aa identity in 229 aa overlap (BLASTp).

```
ORF11  2   LYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLTIIVK 61
           LYAGP+  S +  ++  L+L  DYG + + A P+FWLL  +H+++GNWGW+IIVLT+++K
60K    324 LYAGPKIQSKLKELSPGLELTVDYGFLWFIAQPIFWLLQHIHSLLGNWGWSIIVLTMLIK 383

ORF11  62  AVLYPLTNASYRSMAKMRAAAPKLQAIKEKYGDDRXXXXXXXXXLYTDEKINPLGGCLPM 121
            + +PL+ ASYRSMA+MRA APKL A+KE++GDDR         LY  EKINPLGGCLP+
60K    384 GLFFPLSAASYRSMARMRAVAPKLAALKERFGDDRQKMSQAMMELYKKEKINPLGGCLPI 443

ORF11  122 LLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTYLNPPPT 181
           L+Q+PVF+ LYW L  SVE+RQAPW+ WITDLS  DP++ILPIIM ATMF Q  LNP P
60K    444 LVQMPVFLALYWVLLESVEMRQAPWILWITDLSIKDPFFILPIIMGATMFIQQRLNPTPP 503

ORF11  182 DPMQAKMMKIMPLVXXXXXXXXXPAGXVLYWVVNNLLTIAQQWHINRSIE 230
           DPMQAK+MK+MP++         PAG VLYWVVNN L+I+QQW+I R IE
60K    504 DPMQAKVMKMMPIIFTFFFLWFPAGLVLYWVVNNCLSISQQWYITRRIE 552
```

<u>Homology with a predicted ORF from *N.meningitidis* (strain A)</u>

ORF11 shows 97.9% identity over a 240aa overlap with an ORF (ORF11a) from strain A of *N. meningitidis:*

```
                                                   10        20        30
orf11.pep                               NLYAGPQTTSVIANIADNLQLAKDYGKVHW
                                        ||||||||||||||||||||| ||||||||
orf11a    IKRRNDKLYSTSVSVPLAAIQNGAKSXASINLYAGPQTTSVIANIADNLQLXKDYGKVHW
            280       290       300       310       320       330

                   40        50        60        70        80        90
orf11.pep FASPLFWLLNQLHNIIGNWGWAIIVLTIIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKE
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf11a    FASPLFWLLNQLHNIIGNWGWAIIVLTIIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKE
            340       350       360       370       380       390
```

```
                   100       110       120       130       140       150
orf11.pep  KYGDDRMAQQQAMMQLYTDEKINPLGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWI
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf11a     KYGDDRMAQQQAMMQLYTDEKINPLGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWI
             400       410       420       430       440       450

                   160       170       180       190       200       210
orf11.pep  TDLSRADPYYILPIIMAATMFAQTYLNPPPTDPMQAKMMKIMPLVFSXXFFFFPAGXVLY
           |||||||||||||||||||||||||||||||||||||||||||| ||||| |||| |||
orf11a     TDLSRADPYYILPIIMAATMFAQTYLNPPPTDPMQAKMMKIMPLVXSXXFFXFPAGLVLY
             460       470       480       490       500       510

                   220       230       240
orf11.pep  WVVNNLLTIAQQWHINRSIEKQRAQGEVVSX
           ||:||||||||||||||||||||||||||||
orf11a     WVINNLLTIAQQWHINRSIEKQRAQGEVVSX
             520       530       540
```

The complete length ORF11a nucleotide sequence <SEQ ID 53> is:

```
   1  ANGGATTTTA AAAGACTCAC NGNGTTTTTC GCCATCGCAC TGGTGATTAT
  51  GATCGGATNG NAAANGATGT TCCCCACTCC GAAGCCCGTC CCCGCGCCCC
 101  AACAGACGGC ACAACAACAG GCCGTAANCG CTTCCGCCGA AGCCGCGCTC
 151  GCGCCCGNAN CGCCGATTAC CGTAACGACC GACACGGTTC AAGCCGTCAT
 201  TGATGAAAAA AGCGGCGACC TGCGCCGGCT GACCCTGCTC AAATACAAAG
 251  CAACCGGCGA CNAAAATAAA CCGTTCATCC TGTTTGGCGA CGGCAAANAA
 301  TACACCTACN TCGCCCANTC CGAACTTTTG GACGCGCAGG GCAACAACAT
 351  TCTAAAAGGC ATCGGCTTTA GCGCACCGAA AAAACAGTAC AGCTTGGAAG
 401  GCGACAAAGT TGAAGTCCGC CTGAGCGCAC CTGAAACACG CGGTCTGAAA
 451  ATCGACAAAG TTTATACTTT CACCAAAGGC AGCTATCTGG TCAACGTCCG
 501  CTTCGACATC GCCAACGGCA GCGGTCAAAC CGCCAACCTG AGCGCGGACT
 551  ACCGCATCGT CCGCGACCAC AGCGAACCCG AGGGTCAAGG CTACTTTACC
 601  CACTCTTACG TCGGCCCTGT TGTTTATACC CCTGAAGGCA ACTTCCAAAA
 651  AGTCAGCTTC TCCGACTTGG ACGACGATGC CAANTCCGGN AAATCCGAGG
 701  CCGAATACAT CCGCAAAACC CNGACCGGCT GGCTCGGCAT GATTGAACAC
 751  CACTTCATGT CCACCTGGAT CCTCCAACCC AAAGGCGGAC AAAGCGTTTG
 801  CGCCGCTGGC GACTGCNGTA TNGACATCAA ACGCCGCAAC GACAAGCTGT
 851  ACAGCACCAG CGTCAGCGTG CCTTTAGCCG CTATCCAAAA CGGTGCGAAA
 901  TCCNAAGCCT CCATCAACCT CTACGCCGGC CCACAGACCA CATCNGTTAT
 951  CGCAAACATC GCCGACAACC TGCAACTGGN CAAAGACTAC GGCAAAGTAC
1001  ACTGGTTCGC CTCCCCCCTC TTTTGGCTTT TGAACCAACT GCACAACATC
1051  ATCGGCAACT GGGGCTGGGC GATTATCGTT TTAACCATCA TCGTCAAAGC
1101  CGTACTGTAT CCATTGACCA ACGCCTCTTA CCGTTCGATG GCGAAAATGC
1151  GTGCCGCCGC GCCCAAACTG CAAGCCATCA AAGAGAAATA CGGCGACGAC
1201  CGTATGGCGC AGCAACAAGC CATGATGCAG CTTTACACAG ACGAGAAAAT
1251  CAACCCGCTG GGCGGCTGCC TGCCTATGCT GTTGCAAATC CCCGTCTTCA
1301  TCGGATTGTA TTGGGCATTG TTCGCCTCCG TAGAATTGCG CCAGGCACCT
1351  TGGCTGGGTT GGATTACCGA CCTCAGCCGC GCCGACCCNT ACTACATCCT
1401  GCCCATCATT ATGGCGGCAA CGATGTTCGC CCAAACCTAT CTGAACCCGC
1451  CGCCGACCGA CCCGATGCAG GCGAAAATGA TGAAAATCAT GCCTTTGGTT
1501  NTNTCNNNNA NGTTCTTCNN CTTCCCTGCC GGTCTGGTAT TGTACTGGGT
1551  GATCAACAAC CTCCTGACCA TCGCCCAGCA ATGGCACATC AACCGCAGCA
1601  TCGAAAAACA ACGCGCCCAA GGCGAAGTCG TTTCCTAA
```

This encodes a protein having amino acid sequence <SEQ ID 54>:

```
   1  XDFKRLTXFF AIALVIMIGX XXMFPTPKPV PAPQQTAQQQ AVXASAEAAL
  51  APXXPITVTT DTVQAVIDEK SGDLRRLTLL KYKATGDXNK PFILFGDGKX
 101  YTYXAXSELL DAQGNNILKG IGFSAPKKQY SLEGDKVEVR LSAPETRGLK
 151  IDKVYTFTKG SYLVNVRFDI ANGSGQTANL SADYRIVRDH SEPEGQGYFT
 201  HSYVGPVVYT PEGNFQKVSF SDLDDDAXSG KSEAEYIRKT XTGWLGMIEH
 251  HFMSTWILQP KGGQSVCAAG DCXXDIKRRN DKLYSTSVSV PLAAIQNGAK
 301  SXASINLYAG PQTTSVIANI ADNLQLXKDY GKVHWFASPL FWLLNQLHNI
 351  IGNWGWAIIV LTIIVKAVLY PLTNASYRSM AKMRAAAPKL QAIKEKYGDD
 401  RMAQQQAMMQ LYTDEKINPL GGCLPMLLQI PVFIGLYWAL FASVELRQAP
 451  WLGWITDLSR ADPYYILPII MAATMFAQTY LNPPPTDPMQ AKMMKIMPLV
 501  XSXXFFXFPA GLVLYWVINN LLTIAQQWHI NRSIEKQRAQ GEVVS*
```

ORF11a and ORF11-1 show 95.2% identity in 544 aa overlap:

```
                   10        20        30        40        50        60
```

```
orf11a.pep    XDFKRLTXFFAIALVIMIGXXXMFPTPKPVPAPQQTAQQQAVXASAEAALAPXXPITVTT
               |||||| |||||||||||   |||||||||||||:||||||:||||||||| :||||||
orf11-1       MDFKRLTAFFAIALVIMIGWEKMFPTPKPVPAPQQAAQQQAVTASAEAALAPATPITVTT
                      10        20        30        40        50        60

                      70        80        90       100       110       120
orf11a.pep    DTVQAVIDEKSGDLRRLTLLKYKATGDXNKPFILFGDGKXYTYXAXSELLDAQGNNILKG
              ||||||||||||||||||||||||||| ||||||||||| ||| | ||||||||||||||
orf11-1       DTVQAVIDEKSGDLRRLTLLKYKATGDENKPFILFGDGKEYTYVAQSELLDAQGNNILKG
                      70        80        90       100       110       120

                     130       140       150       160       170       180
orf11a.pep    IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYTFTKGSYLVNVRFDIANGSGQTANL
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf11-1       IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYTFTKGSYLVNVRFDIANGSGQTANL
                     130       140       150       160       170       180

                     190       200       210       220       230       240
orf11a.pep    SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDDAXSGKSEAEYIRKT
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
orf11-1       SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDDAKSGKSEAEYIRKT
                     190       200       210       220       230       240

                     250       260       270       280       290       300
orf11a.pep    XTGWLGMIEHHFMSTWILQPKGGQSVCAAGDCXXDIKRRNDKLYSTSVSVPLAAIQNGAK
               ||||||||||||||||||||| |||||||:|  ||||||||||||||||||||||||||
orf11-1       PTGWLGMIEHHFMSTWILQPKGRQSVCAAGECNIDIKRRNDKLYSTSVSVPLAAIQNGAK
                     250       260       270       280       290       300

                     310       320       330       340       350       360
orf11a.pep    SXASINLYAGPQTTSVIANIADNLQLXKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV
              : |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
orf11-1       AEASINLYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIV
                     310       320       330       340       350       360

                     370       380       390       400       410       420
orf11a.pep    LTIIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKEKYGDDRMAQQQAMMQLYTDEKINPL
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf11-1       LTIIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKEKYGDDRMAQQQAMMQLYTDEKINPL
                     370       380       390       400       410       420

                     430       440       450       460       470       480
orf11a.pep    GGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTY
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf11-1       GGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTY
                     430       440       450       460       470       480

                     490       500       510       520       530       540
orf11a.pep    LNPPPTDPMQAKMMKIMPLVXSXXFFXFPAGLVLYWVINNLLTIAQQWHINRSIEKQRAQ
              |||||||||||||||||||| |  || ||||||||||:||||||||||||||||||||||
orf11-1       LNPPPTDPMQAKMMKIMPLVFSVMFFFFPAGLVLYWVVNNLLTIAQQWHINRSIEKQRAQ
                     490       500       510       520       530       540

orf11a.pep    GEVVSX
              ||||||
orf11-1       GEVVSX
```

<u>Homology with a predicted ORF from *N.gonorrhoeae*</u>

ORF11 shows 96.3% identity over a 240aa overlap with a predicted ORF (ORF11.ng) from *N. gonorrhoeae*:

```
Orf11            NLYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIIVLT   57
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
orf11ng       MAVNLYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIVVLT   60
```

```
orf11      IIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKEKYGDDRMAQQQAMMQLYTDEKINPLGG  117
           ||||||||||||||||||||||||||:||:|||||||||||||||||||: ||:||||||
orf11ng    IIVKAVLYPLTNASYRSMAKMRAAAPELQTIKEKYGDDRMAQQQAMMQLFEDEEINPLGG  120

orf11      CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTYLN  177
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf11ng    CLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTYLN  180

orf11      PPPTDPMQAKMMKIMPLVFSXXFFFFPAGXVLYWVVNNLLTIAQQWHINRSIEKQRAQGE  237
           ||||||||||||||||||||  ||||||| ||||||||||||||||||||||||||||||
orf11ng    PPPTDPMQAKMMKIMPLVFSVMFFFFPAGLVLYWVVNNLLTIAQQWHINRSIEKQRAQGE  240

orf11      VVS  240
           |||
orf11ng    VVS  243
```

An ORF11ng nucleotide sequence <SEQ ID 55> was predicted to encode a protein having amino acid sequence <SEQ ID 56>:

```
  1  MAVNLYAGPQ TTSVIANIAD NLQLAKDYGK VHWFASPLFW LLNQLHNIIG
 51  NWGWAIVVLT IIVKAVLYPL TNASYRSMAK MRAAAPELQT IKEKYGDDRM
101  AQQQAMMQLF EDEEINPLGG CLPMLLQIPV FIGLYWALFA SVELRQAPWL
151  GWITDLSRAD PYYILPIIMA ATMFAQTYLN PPPTDPMQAK MMKIMPLVFS
201  VMFFFFPAGL VLYWVVNNLL TIAQQWHINR SIEKQRAQGE VVS*
```

Further sequence analysis revealed the complete gonococcal DNA sequence <SEQ ID 57> to be:

```
   1  ATGGATTTTA AAAGACTCAC GGCGTTTTTC GCCATCGCGC TGGTGATTAT
  51  GATCGGCTGG GAAAAAATGT TCCCCACCCC GAAACCCGTC CCCGCGCCCC
 101  AACAGGCGGC ACAAAAACAG GCAGCAACCG CTTCCGCCGA AGCCGCGCTC
 151  GCGCCCGCAA CGCCGATTAC CGTAACGACC GACACGGTTC AAGCCGTTAT
 201  TGATGAAAAA AGTGGCGACC TGCGCCGGCT GACCCTGCTC AAATACAAAG
 251  CAACCGGCGA CGAAAACAAA CCGTTCGTCC TGTTTGGCGA CGGCAAAGAA
 301  TACACCTACG TCGCCCAATC CGAACTTTTG GACGCGCAGG GCAACAACAT
 351  TCTGAAAGGC ATCGGCTTTA GCGCACCGAA AAAACAGTAC ACCCTCAACG
 401  GCGACACAGT CGAAGTCCGC CTGAGCGCGC CCGAAACCAA CGGACTGAAA
 451  ATCGACAAAG TCTATACCTT TACCAAAGAC AGCTATCTGG TCAACGTCCG
 501  CTTCGACATC GCCAACGGCA GCGGTCAAAC CGCCAACCTG AGCGCGGACT
 551  ACCGCATCGT CCGCGACCAC AGCGAACCCG AGGGTCAAGG CTACTTTACC
 601  CACTCTTACG TCGGCCCTGT TGTTTATACC CCTGAAGGCA ACTTCCAAAA
 651  AGTCAGCTTC TCCgacTTgg acgACGATGC gaaaTccggc aaATccgagg
 701  ccgaatacaT CCGCAAAACC ccgaccggtt ggctcggcat gattgaacac
 751  cacttcatgt ccacctggat cctccAAcct aaaggcggcc aaaacgtttg
 801  cgcccaggga gactgccgta tcgacattaa aCgccgcaac gacaagctgt
 851  acagcgcaag cgtcagcgtg cctttaaccg ctatcccaac ccgggggcca
 901  aaaccgaaaa tggcggTCAA CCTGTATGCC GGTCCGCAAA CCACATCCGT
 951  TATCGCAAAC ATCGCCgacA ACCTGCAACT GGCAAAAGAC TACGGTAAAG
1001  TACACTGGTT CGCATCGCCG CTCTTCTGGC TCCTGAACCA ACTGCACAAC
1051  ATTATCGGCA ACTGGGGCTG GGCAATCGTC GTTTTGACCA TCATCGTCAA
1101  AGCCGTACTG TATCCATTGA CCAACGcctc ctACCGTTCG ATGGCGAAAA
1151  TGCGTGccgc cgcacCcaaA CTGCAGACCA TCAAAGAAAA ATAcgGCGAC
1201  GACCGTATGG CGCAACAGCA AGCGATGATG CAGCTTTACA AAgacgAGAA
1251  AATCAACCCG CTGGGCGGCT GTctgcctat gctgttgCAA ATCCCCGTCT
1301  TCATCGGCTT GTACTGGGCA TTGTTCGCCT CCGTAGAATT GCGCCAGGCA
1351  CCTTGGCTGG GCTGGATTAC CGACCTCAGC CGCGCCGACC CCTACTACAT
1401  CCTGCCCATC ATTATGGCGG CAACGATGTT CGCCCAAACC TATCTGAACC
1451  CGCCGCCGAC CGACCCGATG CAGGCGAAAA TGATGAAAAT CATGCCGTTG
1501  GTTTTCTCCG TCATGTTCTT CTTCTTCCCT GCCGGTTTGG TTCTCTACTG
1551  GGTGGTCAAC AACCTCCTGA CCATCGCCCA GCAGTGGCAC ATCAACCGCA
1601  GCATCGAAAA ACAACGCGCC CAAGGCGAAG TCGTTTCCTA A
```

This encodes a protein having amino acid sequence <SEQ ID 58; ORF11ng-1>:

```
  1  MDFKRLTAFF AIALVIMIGW EKMFPTPKPV PAPQQAAQKQ AATASAEAAL
 51  APATPITVTT DTVQAVIDEK SGDLRRLTLL KYKATGDENK PFVLFGDGKE
101  YTYVAQSELL DAQGNNILKG IGFSAPKKQY TLNGDTVEVR LSAPETNGLK
151  IDKVYTFTKD SYLVNVRFDI ANGSGQTANL SADYRIVRDH SEPEGQGYFT
201  HSYVGPVVYT PEGNFQKVSF SDLDDDAKSG KSEAEYIRKT PTGWLGMIEH
251  HFMSTWILQP KGGQNVCAQG DCRIDIKRRN DKLYSASVSV PLTAIPTRGP
301  KPKMAVNLYA GPQTTSVIAN IADNLQLAKD YGKVHWFASP LFWLLNQLHN
```

```
351  IIGNWGWAIV VLTIIVKAVL YPLTNASYRS MAKMRAAAPK LQTIKEKYGD
401  DRMAQQQAMM QLYKDEKINP LGGCLPMLLQ IPVFIGLYWA LFASVELRQA
451  PWLGWITDLS RADPYYILPI IMAATMFAQT YLNPPPTDPM QAKMMKIMPL
501  VFSVMFFFFP AGLVLYWVVN NLLTIAQQWH INRSIEKQRA QGEVVS*
```

ORF11ng-1 and ORF11-1 shown 95.1% identity in 546 aa overlap:

```
                           10        20        30        40        50        60
   orf11ng-1.pep  MDFKRLTAFFAIALVIMIGWEKMFPTPKPVPAPQQAAQKQAATASAEAALAPATPITVTT
                  |||||||||||||||||||||||||||||||||||||||:||:|||||||||||||||||
   orf11-1        MDFKRLTAFFAIALVIMIGWEKMFPTPKPVPAPQQAAQQQAVTASAEAALAPATPITVTT
                           10        20        30        40        50        60

                           70        80        90       100       110       120
   orf11ng-1.pep  DTVQAVIDEKSGDLRRLTLLKYKATGDENKPFVLFGDGKEYTYVAQSELLDAQGNNILKG
                  |||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
   orf11-1        DTVQAVIDEKSGDLRRLTLLKYKATGDENKPFILFGDGKEYTYVAQSELLDAQGNNILKG
                           70        80        90       100       110       120

                          130       140       150       160       170       180
   orf11ng-1.pep  IGFSAPKKQYTLNGDTVEVRLSAPETNGLKIDKVYTFTKDSYLVNVRFDIANGSGQTANL
                  ||||||||||:|:|| |||||||||| |||||||||||| ||||||||||||||||||||
   orf11-1        IGFSAPKKQYSLEGDKVEVRLSAPETRGLKIDKVYTFTKGSYLVNVRFDIANGSGQTANL
                          130       140       150       160       170       180

                          190       200       210       220       230       240
   orf11ng-1.pep  SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDDAKSGKSEAEYIRKT
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   orf11-1        SADYRIVRDHSEPEGQGYFTHSYVGPVVYTPEGNFQKVSFSDLDDDAKSGKSEAEYIRKT
                          190       200       210       220       230       240

                          250       260       270       280       290       300
   orf11ng-1.pep  PTGWLGMIEHHFMSTWILQPKGGQNVCAQGDCRIDIKRRNDKLYSASVSVPLTAIPTRGP
                  |||||||||||||||||||||| |:||| |:| ||||||||||||:||||||:|| : | 
   orf11-1        PTGWLGMIEHHFMSTWILQPKGRQSVCAAGECNIDIKRRNDKLYSTSVSVPLAAIQN-GA
                          250       260       270       280       290

                          310       320       330       340       350       360
   orf11ng-1.pep  KPKMAVNLYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAIV
                  | : ::|||||||||||||||||||||||||||||||||||||||||||||||||||||:
   orf11-1        KAEASINLYAGPQTTSVIANIADNLQLAKDYGKVHWFASPLFWLLNQLHNIIGNWGWAII
                300       310       320       330       340       350

                          370       380       390       400       410       420
   orf11ng-1.pep  VLTIIVKAVLYPLTNASYRSMAKMRAAAPKLQTIKEKYGDDRMAQQQAMMQLYKDEKINP
                  ||||||||||||||||||||||||||||||||:|||||||||||||||||||| ||||||
   orf11-1        VLTIIVKAVLYPLTNASYRSMAKMRAAAPKLQAIKEKYGDDRMAQQQAMMQLYTDEKINP
                360       370       380       390       400       410

                          430       440       450       460       470       480
   orf11ng-1.pep  LGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQT
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   orf11-1        LGGCLPMLLQIPVFIGLYWALFASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQT
                420       430       440       450       460       470

                          490       500       510       520       530       540
   orf11ng-1.pep  YLNPPPTDPMQAKMMKIMPLVFSVMFFFFPAGLVLYWVVNNLLTIAQQWHINRSIEKQRA
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   orf11-1        YLNPPPTDPMQAKMMKIMPLVFSVMFFFFPAGLVLYWVVNNLLTIAQQWHINRSIEKQRA
                480       490       500       510       520       530

   orf11ng-1.pep  QGEVVSX
                  |||||||
   orf11-1        QGEVVSX
                540
```

In addition, ORF11ng-1 shows significant homology with an inner-membrane protein from the database (accession number p25754):

```
ID   60IM_PSEPU     STANDARD;      PRT;   560 AA.
AC   P25754;
DT   01-MAY-1992 (REL. 22, CREATED)
DT   01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT   01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE   60 KD INNER-MEMBRANE PROTEIN. . . .

SCORES      Init1:  1074 Initn:  1293 Opt:  1103
Smith-Waterman score: 1406;    41.5% identity in 574 aa overlap

                             10        20                            30        40
orf1ng-1.pep    MDFKR---LTAFFAIALVIMIGW-----EKMFPT------------PKPVPAPQQAAQKQ
                ||:||   ::|: ::: |::: |      :  :||            |  ||| :::|: :
p25754          MDIKRTILIAALAVVSYVMVLKWNDDYGQQAALPTQNTAASTVAPGLPDGVPAGNNGASAD
                         10        20        30        40        50        60

                     50               60        70        80        90
orf1ng-1.pep    AATASAEAALAPATPIT-------VTTDTVQAVIDEKSGDLRRLTLLKYKATGDE-NKPF
                : :|:||:: |  :|::       | ||::: :||  :||: :|:| ||    |: | ||
p25754          VPSANAESSPAELAPVALSKDLIRVKTDVLELAIDPVGGDIVQLNLPKYPRRQDHPNIPF
                         70        80        90       100       110       120

                    100       110       120          130           140
orf1ng-1.pep    VLFGDGKEYTYVAQSELLDAQGNNILKGIG---FSAPKKQYTL-NGD---TVEVRLSAPE
                || :| | :|:||| |   ::| :   :: |     ::| :|:| | :|:    :|::::|
p25754          QLFDNGGERVYLAQSGLTGTDGPDA-RASGRPLYAAEQKSYQLADGQEQLVVDLKFS---
                        130       140          150       160       170

                  150       160       170       180       190        200
orf1ng-1.pep    TNGLKIDKVYTFTKDSYLVNVRFDIANGSGQTANLSADYRIVRDHS-EPEGQGYF-THSY
                 ||::  | ::| :    | :|| : | |  |||: | :      :: || | :| ::     | :|
p25754          DNGVNYIKRFSFKRGEYDLNVSYLIDNQSGQAWNGNMFAQLKRDASGDPSSSTATGTATY
                 180       190       200       210       220       230

                    210       220       230       240       250       260
orf1ng-1.pep    VGPVVYTPEGNFQKVSFSDLDDDAKSGKSEAEYIRKTPTGWLGMIEHHFMSTWILQPKGG
                :| :::|      ::|||::|:|    |:: :|      ::    ||:: ::|:|:::||    |:
p25754          LGAALWTASEPYKKVSMKDID---KGSLKE-----NVSGGWVAWLQHYFVTAWI-PAKSD
                 240       250          260            270       280

                    270       280       290       300       310       320
orf1ng-1.pep    QNVCAQGDCRIDIKRRNDKLYSASVSVPLTAIPTRGPKPKMAVNLYAGPQTTSVIANIAD
                :||       :: :: ::   |   :  :  |: ::|: |  | : ::  |||||:   | : :::
p25754          NNV-------VQTRKDSQGNYIIGYTGPVISVPA-GGKVETSALLYAGPKIQSKLKELSP
                290                300       310        320       330

                    330       340       350       360       370       380
orf1ng-1.pep    NLQLAKDYGKVHWF-ASPLFWLLNQLHNIIGNWGWAIVVLTIIVKAVLYPLTNASYRSMA
                :|:|: |||  :  || |:|:||||:::|:::|||||:|:||||:::|::::||: ||||||
p25754          GLELTVDYGFL-WFIAQPIFWLLQHIHSLLGNWGWSIIVLTMLIKGLFFPLSAASYRSMA
              340       350       360       370       380       390

                    390       400       410       420       430       440
orf1ng-1.pep    KMRAAAPKLQTIKEKYGDDRMAQQQAMMQLYKDEKINPLGGCLPMLLQIPVFIGLYWALF
                :|||:|||| ::||::||||: ::||||:||| |||||||||||:|:|:|||::|||:|:
p25754          RMRAVAPKLAALKERFGDDRQKMSQAMMELYKKEKINPLGGCLPILVQMPVFLALYWVLL
               400       410       420       430       440       450

                    450       460       470       480       490       500
orf1ng-1.pep    ASVELRQAPWLGWITDLSRADPYYILPIIMAATMFAQTYLNPPPTDPMQAKMMKIMPLVF
                 |||:|||||: ||||| ||::|||||||:|||| |   ||| | ||||||:||:||::|
p25754          ESVEMRQAPWILWITDLSIKDPFFILPIIMGATMFIQQRLNPTPPDPMQAKVMKMMPIIF
               460       470       480       490       500       510

                     510       520       530       540
orf1ng-1.pep    SVMFFFFPAGLVLYWVVNNLLTIAQQWHINRSIEKQRAQGEVVSX
                : :|::|||||||||||||| |:|:|||:|:| ||
p25754          TFFFLWFPAGLVLYWVVNNCLSISQQWYITRRIEAATKKAAA
               520       530       540       550       560
```

Based on this analysis, including the homology to an inner-membrane protein from *P. putida* and the predicted transmembrane domains (seen in both the meningococcal and gonoccal proteins), it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 8

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 59>:

```
  1  ..GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGGTTGT
 51    NAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
101    CGCCTGCCGC CGTCTTGACC GNCGCTCTGC TTTCCGCGCT GGGTATTTNG
151    TTCGTACACG CCAAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
201    GGATTTGGAT GCCGGACAAT ATGTCGAAAT CCTCCGNCAC ACAGGCGGCA
251    ACCGTTACGA AGTT.TTTAT CGCGGTACG. ACTGGCAGGC TCAAAATACG
301    GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATTG TCCGCAAGGA
351    AGGCAACCTT CTTATTATCA CACACCCTTA A
```

This corresponds to the amino acid sequence <SEQ ID 60; ORF13>:

```
  1  ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
 51    FVHAKTAVRK VETDSYQDLD AGQYVEILRH TGGNRYEVXY RGTXWQAQNT
101    GQEELEPGTR ALIVRKEGNL LIITHP*
```

Further sequence analysis elaborated the DNA sequence slightly <SEQ ID 61>:

```
  1  ..GCCGTCTTAA TCATCGAATT ATTGACGGGA ACGGTTTATC TTTTGGTTGT
 51    nAGCGCGGCT TTGGCGGGTT CGGGCATTGC TTACGGGCTG ACCGGCAGTA
101    CGCCTGCCGC CGTCTTGACC GnCGCTCTGC TTTCCGCGCT GGGTATTTnG
151    TTCGTACACG CCAAAACCGC CGTTAGAAAA GTTGAAACGG ATTCATATCA
201    GGATTTGGAT GCCGGACAAT ATGTCGAAAT CCTCCGACAC ACAGGCGGCA
251    ACCGTTACGA AGTTTTtTAT CGCGGTACGc ACTGGCAGGC TCAAAATACG
301    GGGCAAGAAG AGCTTGAACC AGGAACTCGC GCCCTCATTG TCCGCAAGGA
351    AGGCAACCTT CTTATTATCA CACACCCTTA A
```

This corresponds to the amino acid sequence <SEQ ID 62; ORF13-1>:

```
  1  ..AVLIIELLTG TVYLLVVSAA LAGSGIAYGL TGSTPAAVLT XALLSALGIX
 51    FVHAKTAVRK VETDSYQDLD AGQYVEILRH TGGNRYEVFY RGTHWQAQNT
101    GQEELEPGTR ALIVRKEGNL LIITHP*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF13 shows 92.9% identity over a 126aa overlap with an ORF (ORF13a) from strain A of *N. meningitidis:*

```
                            10        20        30        40        50
orf13.pep          AVLIIELLTGTVYLLVVSAALAGSGIAYGLTGSTPAAVLTXALLSALGIXF
                   |||||||||||||||||||||||||||||||||||||||| |||||||| |
orf13a    MTVWFVAAVAVLIIELLTGTVYLLVVSAALAGSGIAYGLTGSTPAAVLTAALLSALGIWF
                  10        20        30        40        50        60

                 60        70        80        90       100       110
orf13.pep VHAKTAVRKVETDSYQDLDAGQYVEILRHTGGNRYEVXYRGTXWQAQNTGQEELEPGTRA
          ||||||| |||||||||||||||:|||||:||||||| |||| |||||||||||||||||
orf13a    VHAKTAVGKVETDSYQDLDAGQYAEILRHAGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
                  70        80        90       100       110       120

                120
orf13.pep LIVRKEGNLLIITHPX
          ||||||||||||::||
```

```
orf13a        LIVRKEGNLLIIAKPX
                     130
```

The complete length ORF13a nucleotide sequence <SEQ ID 63> is:

```
  1  ATGACTGTAT GGTTTGTTGC CGCTGTTGCC GTCTTAATCA TCGAATTATT
 51  GACGGGAACG GTTTATCTTT TGGTTGTCAG CGCGGCTTTG GCGGGTTCGG
101  GCATTGCTTA CGGGCTGACC GGCAGCACGC CTGCCGCCGT CTTGACCGCC
151  GCTCTGCTTT CCGCGCTGGG TATTTGGTTC GTACACGCCA AAACCGCCGT
201  GGGAAAAGTT GAAACGGATT CATATCAGGA TTTGGATGCC GGGCAATATG
251  CCGAAATCCT CCGGCACGCA GGCGGCAACC GTTACGAAGT TTTTTATCGC
301  GGTACGCACT GGCAGGCTCA AAATACGGGG CAAGAAGAGC TTGAACCAGG
351  AACGCGCGCC CTAATCGTCC GCAAGGAAGG CAACCTTCTT ATCATCGCAA
401  AACCTTAA
```

This encodes a protein having amino acid sequence <SEQ ID 64>:

```
  1  MTVWFVAAVA VLIIELLTGT VYLLVVSAAL AGSGIAYGLT GSTPAAVLTA
 51  ALLSALGIWF VHAKTAVGKV ETDSYQDLDA GQYAEILRHA GGNRYEVFYR
101  GTHWQAQNTG QEELEPGTRA LIVRKEGNLL IIAKP*
```

ORF13a and ORF13-1 show 94.4% identity in 126 aa overlap

```
                        10        20        30        40        50        60
orf13a.pep    MTVWFVAAVAVLIIELLTGTVYLLVVSAALAGSGIAYGLTGSTPAAVLTAALLSALGIWF
                        |||||||||||||||||||||||||||||||||||||||| |||||||| |
orf13-1                 AVLIIELLTGTVYLLVVSAALAGSGIAYGLTGSTPAAVLTXALLSALGIXF
                                10        20        30        40        50

                        70        80        90       100       110       120
orf13a.pep    VHAKTAVGKVETDSYQDLDAGQYAEILRHAGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
              ||||||| |||||||||||||||:|||||:||||||||||||||||||||||||||||||
orf13-1       VHAKTAVRKVETDSYQDLDAGQYVEILRHTGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
                      60        70        80        90       100       110

                     130
orf13a.pep    LIVRKEGNLLIIAKPX
              ||||||||||||::||
orf13-1       LIVRKEGNLLIITHPX
                    120
```

<u>Homology with a predicted ORF from *N.gonorrhoeae*</u>

ORF13 shows 89.7% identity over a 126aa overlap with a predicted ORF (ORF13.ng) from *N. gonorrhoeae*:

```
orf13                  AVLIIELLTGTVYLLVVSAALAGSGIAYGLTGSTPAAVLTXALLSALGIXF   51
                       |||||||||||||||||||||||||||||||||||||||| |||||||| |
orf13ng       MTVWFVAAVAVLIIELLTGTVYLLVVSAALAGSGIAYGLTGSTPAAVLTAALLSALGIWF   60

orf13         VHAKTAVRKVETDSYQDLDAGQYVEILRHTGGNRYEVXYRGTXWQAQNTGQEELEPGTRA  111
              ||||||| |||||||||||:|:|:||||:|||||||| |||| ||||||||| :||||||
orf13ng       VHAKTAVGKVETDSYQDLDTGKYAEILRYTGGNRYEVFYRGTHWQAQNTGQEVFEPGTRA  120

orf13         LIVRKEGNLLIITHP  126
              ||||||||||||::|
orf13ng       LIVRKEGNLLIIANP  135
```

The complete length ORF13ng nucleotide sequence <SEQ ID 65> is:

```
  1  ATGACTGTAT GGTTTGTTGC CGCTGTTGCC GTCTTAATCA TCGAATTATT
 51  GACGGGAACG GTTTATCTTT TGGTTGTCAG CGCGGCTTTG GCGGGTTCGG
101  GCATTGCCTA CGGGCTGACT GGCAGCACGC CTGCCGCCGT CTTGACCGCC
151  GCACTGCTTT CCGCGCTGGG CATTTGGTTC GTACATGCCA AAACCGCCGT
201  GGGAAAAGTT GAAACGGATT CATATCAGGA TTTGGATACC GGAAAATATG
251  CCGAAATCCT CCGATACACA GGCGGCAACC GTTACGAAGT TTTTTATCGC
301  GGTACGCACT GGCAGGCGCA AAATACGGGG CAGGAAGTGT TTGAACCGGG
351  AACGCGCGCC CTCATCGTCC GCAAAGAAGG TAACCTTCTT ATCATCGCAA
401  ACCCTTAA
```

This encodes a protein having amino acid sequence <SEQ ID 66>:

```
  1  MTVWFVAAVA VLIIELLTGT VYLLVVSAAL AGSGIAYGLT GSTPAAVLTA
 51  ALLSALGIWF VHAKTAVGKV ETDSYQDLDT GKYAEILRYT GGNRYEVFYR
101  GTHWQAQNTG QEVFEPGTRA LIVRKEGNLL IIANP*
```

ORF13ng shows 91.3% identity in 126 aa overlap with ORF13-1:

```
                                     10        20        30        40        50
orf13-1.pep                  AVLIIELLTGTVYLLVVSAALAGSGIAYGLTGSTPAAVLTXALLSALGIXF
                             |||||||||||||||||||||||||||||||||||||||| |||||||| |
orf13ng             MTVWFVAAVAVLIIELLTGTVYLLVVSAALAGSGIAYGLTGSTPAAVLTAALLSALGIWF
                            10        20        30        40        50        60

                           60        70        80        90       100       110
orf13-1.pep         VHAKTAVRKVETDSYQDLDAGQYVEILRHTGGNRYEVFYRGTHWQAQNTGQEELEPGTRA
                    ||||||| |||||||||||:|:|:||||:||||||||||||||||||||||| :||||||
orf13ng             VHAKTAVGKVETDSYQDLDTGKYAEILRYTGGNRYEVFYRGTHWQAQNTGQEVFEPGTRA
                            70        80        90       100       110       120

                          120
orf13-1.pep         LIVRKEGNLLIITHPX
                    ||||||||||||::||
orf13ng             LIVRKEGNLLIIANPX
                           130
```

Based on this analysis, including the extensive leader sequence in this protein, it is predicted that ORF13 and ORF13ng are likely to be outer membrane proteins. It is thus predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 9

The following DNA sequence was identified in *N.meningitidis* <SEQ ID 67>:

```
  1  ATGTwTGATT TCGGTTTrGG CGArCTGGTT TTTGTCGGCA TTATCGCCCT
 51  GATwGtCCTC GGCCCCGAAC GCsTGCCCGA GGCCGCCCGC AyCGCCGGAC
101  GGcTCATCGG CAGGCTGCAA CGCTTTGTCG GcAGCGTCAA ACAGGAATTT
151  GACACTCAAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
201  AGCTGCCGcC GCTCAGGTTC GAGACAGCCT CAAAGAAACC GGTACGGATA
251  TGGAAGGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
301  CTGCCCGAAC AGCGGACACC TGCCGATTTC GGTGTCGATG AAAACGGCAA
351  TCCGCT.TCC CGATGCGGCA AACACCCTAT CAGACGGCAT TTCCGACGTT
401  ATGCCGTC..
```

This corresponds to the amino acid sequence <SEQ ID 68; ORF2>:

```
  1  MXDFGLGELV FVGIIALIVL GPERXPEAAR XAGRLIGRLQ RFVGSVKQEF
 51  DTQIELEELR KAKQEFEAAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
101  LPEQRTPADF GVDENGNPXS RCGKHPIRRH FRRYAV..
```

Further work revealed the complete nucleotide sequence <SEQ ID 69>:

```
  1  ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTCGGCA TTATCGCCCT
 51  GATTGTCCTC GGCCCCGAAC GCCTGCCCGA GGCCGCCCGC ACCGCCGGAC
101  GGCTCATCGG CAGGCTGCAA CGCTTTGTCG GCAGCGTCAA ACAGGAATTT
151  GACACTCAAA TCGAACTGGA AGAACTGAGG AAGGCAAAGC AGGAATTTGA
201  AGCTGCCGCC GCTCAGGTTC GAGACAGCCT CAAAGAAACC GGTACGGATA
251  TGGAAGGCAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
301  CTGCCCGAAC AGCGGACACC TGCCGATTTC GGTGTCGATG AAAACGGCAA
351  TCCGCTTCCC GATGCGGCAA ACACCCTATC AGACGGCATT TCCGACGTTA
401  TGCCGTCCGA ACGTTCCTAC GCTTCCGCCG AAACCCTTGG GGACAGCGGG
451  CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGCGCATG
501  GCGGGAATAC CTGACTGCTT CTGCCGCCGC ACCCGTCGTA CAGACCGTCG
```

```
551  AAGTCAGCTA TATCGATACT GCTGTTGAAA CGCCTGTTCC GCACACCACT
601  TCCCTGCGCA AACAGGCAAT AAGCCGCAAA CGCGATTTTC GTCCGAAACA
651  CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA
```

This corresponds to the amino acid sequence <SEQ ID 70; ORF2-1>:

```
  1  MFDFGLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF
 51  DTQIELEELR KAKQEFEAAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
101  LPEQRTPADF GVDENGNPLP DAANTLSDGI SDVMPSERSY ASAETLGDSG
151  QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETPVPHTT
201  SLRKQAISRK RDFRPKHRAK PKLRVRKS*
```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 71 >:

```
  1  ATGTTTGATT TCGGTTTGGG CGAGCTGGTT TTTGTCGGCA TTATCGCCCT
 51  GATTGTCCTC GGCCCCGAAC GCCTGCCCGA GGCCGCCCGC ACCGCCGGAC
101  GGCTCATCGG CAGGCTGCAA CGCTTTGTCG GCAGCGTCAA ACAGGAATTT
151  GACACGCAAA TCGAACTGGA AGAACTAAGG AAGGCAAAGC AGGAATTTGA
201  AGCTGCCGCT GCTCAGGTTC GAGACAGCCT CAAAGAAACC GGTACGGATA
251  TGGAGGGTAA TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
301  CTGCCCGAAC AGCGCACGCC TGCTGATTTC GGTGTCGATG AAAACGGCAA
351  TCCCTTTCCC GATGCGGCAA ACACCCTATT AGACGGCATT TCCGACGTTA
401  TGCCGTCCGA ACGTTCCTAC GCTTCCGCCG AAACCCTTGG GGACAGCGGG
451  CAAACCGGCA GTACAGCCGA ACCCGCGGAA ACCGACCAAG ACCGTGCATG
501  GCGGGAATAC CTGACTGCTT CTGCCGCCGC ACCCGTCGTA CAGACCGTCG
551  AAGTCAGCTA TATCGATACC GCTGTTGAAA CCCCTGTTCC GCATACCACT
601  TCGCTGCGTA AACAGGCAAT AAGCCGCAAA CGCGATTTGC GTCCTAAATC
651  CCGCGCCAAA CCTAAATTGC GCGTCCGTAA ATCATAA
```

This encodes a protein having amino acid sequence <SEQ ID 72; ORF2a>:

```
  1  MFDFGLGELV FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEF
 51  DTQIELEELR KAKQEFEAAA AQVRDSLKET GTDMEGNLHD ISDGLKPWEK
101  LPEQRTPADF GVDENGNPFP DAANTLLDGI SDVMPSERSY ASAETLGDSG
151  QTGSTAEPAE TDQDRAWREY LTASAAAPVV QTVEVSYIDT AVETPVPHTT
201  SLRKQAISRK RDLRPKSRAK PKLRVRKS*
```

The originally-identified partial strain B sequence (ORF2) shows 97.5% identity over a 118aa overlap with ORF2a:

```
                          10        20        30        40        50        60
orf2.pep         MXDFGLGELVFVGIIALIVLGPERXPEAARXAGRLIGRLQRFVGSVKQEFDTQIELEELR
                 | |||||||||||||||||||||| |||||:|||||||||||||||||||||||||||||
orf2a            MFDFGLGELVFVGIIALIVLGPERLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR
                          10        20        30        40        50        60

                          70        80        90       100       110       120
orf2.pep         KAKQEFEAAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPXS
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf2a            KAKQEFEAAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPFP
                          70        80        90       100       110       120

                         130
orf2.pep         RCGKHPIRRHFRRYAV

orf2a            DAANTLLDGISDVMPSERSYASAETLGDSGQTGSTAEPAETDQDRAWREYLTASAAAPVV
                         130       140       150       160       170       180
```

The complete strain B sequence (ORF2-1) and ORF2a show 98.2% identity in 228 aa overlap:

```
orf2a.pep    MFDFGLGELVFVGIIALIVLGPERLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf2-1       MFDFGLGELVFVGIIALIVLGPERLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR  60

orf2a.pep    KAKQEFEAAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPFP 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
orf2-1       KAKQEFEAAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPLP 120

orf2a.pep    DAANTLLDGISDVMPSERSYASAETLGDSGQTGSTAEPAETDQDRAWREYLTASAAAPVV 180
             |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
orf2-1        DAANTLSDGISDVMPSERSYASAETLGDSGQTGSTAEPAETDQDRAWREYLTASAAAPVV 180

orf2a.pep     QTVEVSYIDTAVETPVPHTTSLRKQAISRKRDLRPKSRAKPKLRVRKSX 229
              ||||||||||||||||||||||||||||||||:||| |||||||||||
orf2-1        QTVEVSYIDTAVETPVPHTTSLRKQAISRKRDFRPKHRAKPKLRVRKSX 229
```

Further work identified a partial DNA sequence <SEQ ID 73> in *N.gonorrhoeae* encoding the following amino acid sequence <SEQ ID 74; ORF2ng>:

```
  1  MFDFGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL
 51  DTQIELEELR KVKQAFEAAA AQVRDSLKET DTDMQNSLHD ISDGLKPWEK
101  LPEQRTPADF GVDEKGNSLS RYGKHRIRRH FRRYAV*
```

Further work identified the complete gonococcal gene sequence <SEQ ID 75>:

```
  1  ATGTTTGATT TCGGTTTGGG CGAGCTGATT TTTGTCGGCA TTATCGCCCT
 51  GATTGTCCTT GGTCCAGAAC GCCTGCCCGA AGCCGCCCGC ACTGCCGGAC
101  GGCTTATCGG CAGGCTGCAA CGCTTTGTAG GAAGCGTCAA ACAAGAACTT
151  GACACTCAAA TCGAACTGGA AGAGCTGAGG AAGGTCAAGC AGGCATTCGA
201  AGCTGCCGCC GCTCAGGTTC GAGACAGCCT CAAAGAAACC GATACGGATA
251  TGCAGAACAG TCTGCACGAC ATTTCCGACG GTCTGAAGCC TTGGGAAAAA
301  CTGCCCGAAC AGCGCACGCc tgccgatttc gGTGTCGATg AAAacggcaa
351  tcccttccc gATACGGCAA ACACCGTATC AGACGGCATT TCCGACGTTA
401  TGCCGTCTGA ACGTTCCGAT ACTtccgcCG AAACCCTTGG GGACGACAGG
451  CAAACCGGCA GTACAGCCGA ACCTGCGGAA ACCGACAAAG ACCGCGCATG
501  GCGGGAATAC CTGactgctt ctgccgccgc acctgtcgta Cagagggccg
551  tcgaagtcag ctaTATCGAT ACTGCTGTTG AAacgcctgT tccgcaCacc
601  acttccctgc gcaAACAGGC AATAAACCGC AAACGCGATT TttgtccgaA
651  ACACCGCGCc aAACCGAAat tgcgcgtcCG TAAATCATAA
```

This encodes a protein having the amino acid sequence <SEQ ID 76; ORF2ng-1>:

```
  1  MFDFGLGELI FVGIIALIVL GPERLPEAAR TAGRLIGRLQ RFVGSVKQEL
 51  DTQIELEELR KVKQAFEAAA AQVRDSLKET DTDMQNSLHD ISDGLKPWEK
101  LPEQRTPADF GVDENGNPLP DTANTVSDGI SDVMPSERSD TSAETLGDDR
151  QTGSTAEPAE TDKDRAWREY LTASAAAPVV QRAVEVSYID TAVETPVPHT
201  TSLRKQAINR KRDFCPKHRA KPKLRVRKS*
```

The originally-identified partial strain B sequence (ORF2) shows 87.5% identity over a 136aa overlap with ORF2ng:

```
orf2.pep     MXDFGLGELVFVGIIALIVLGPERXPEAARXAGRLIGRLQRFVGSVKQEFDTQIELEELR   60
             | |||||||:|||||||||||||| |||||:||||||||||||||||||:||||||||||
orf2ng       MFDFGLGELIFVGIIALIVLGPERLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR   60

orf2.pep     KAKQEFEAAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPXS  120
             |:|| ||||||||||||||| |||:::|||||||||||||||||||||||||||:||
orf2ng       KVKQAFEAAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLPEQRTPADFGVDEKGNSLP  120

orf2.pep     RCGKHPIRRHFRRYAV  136
             | ||| ||||||||||
orf2ng       RYGKHRIRRHFRRYAV  136
```

The complete strain B and gonococcal sequences (ORF2-1 & ORF2ng-1) show 91.7% identity in 229 aa overlap:

```
                      10        20        30        40        50        60
orf2-1.pep    MFDFGLGELVFVGIIALIVLGPERLPEAARTAGRLIGRLQRFVGSVKQEFDTQIELEELR
              |||||||||:|||||||||||||||||||||||||||||||||||||||:||||||||||
orf2ng-1      MFDFGLGELIFVGIIALIVLGPERLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR
                      10        20        30        40        50        60

                      70        80        90       100       110       120
orf2-1.pep    KAKQEFEAAAAQVRDSLKETGTDMEGNLHDISDGLKPWEKLPEQRTPADFGVDENGNPLP
              |:|| ||||||||||||||| |||:::|||||||||||||||||||||||||||||||||
orf2ng-1      KVKQAFEAAAAQVRDSLKETDTDMQNSLHDISDGLKPWEKLPEQRTPADFGVDENGNPLP
```

```
                           70        80        90       100       110       120

                          130       140       150       160       170       180
orf2-1.pep       DAANTLSDGISDVMPSERSYASAETLGDSGQTGSTAEPAETDQDRAWREYLTASAAAPVV
                 |:|||:||||||||||||| :|||||||: ||||||||||||:|||||||||||||||||
orf2ng-1         DTANTVSDGISDVMPSERSDTSAETLGDDRQTGSTAEPAETDKDRAWREYLTASAAAPVV
                          130       140       150       160       170       180

                           190       200       210       220       229
orf2-1.pep       Q-TVEVSYIDTAVETPVPHTTSLRKQAISRKRDFRPKHRAKPKLRVRKSX
                 | :|||||||||||||||||||||||||:||||| |||||||||||||||
orf2ng-1         QRAVEVSYIDTAVETPVPHTTSLRKQAINRKRDFCPKHRAKPKLRVRKSX
                          190       200       210       220       230
```

Computer analysis of these amino acid sequences indicates a transmembrane region (underlined), and also revealed homology (59% identity) between the gonococcal sequence and the TatB protein of *E.coli*:

```
 gnl|PID|e1292181 (AJ005830) TatB protein [Escherichia coli] Length = 171
 Score = 56.6 bits (134), Expect = 1e-07
 Identities = 30/88 (34%), Positives = 52/88 (59%), Gaps = 1/88 (1%)

Query: 1  MFDFGLGELIFVGIIALIVLGPERLPEAARTAGRLIGRLQRFVGSVKQELDTQIELEELR 60
           MFD G   EL+ V II L+VLGP+RLP A +T     I  L+     +V+ EL  +++L+E +
Sbjct: 1  MFDIGFSELLLVFIIGLVVLGPQRLPVAVKTVAGWIRALRSLATTVQNELTQELKLQEFQ 60

Query: 61 -KVKQAFEAAAAQVRDSLKETDTDMQNS 87
             +K+  +A+    +    LK +  +++ +
Sbjct: 61 DSLKKVEKASLTNLTPELKASMDELRQA 88
```

Based on this analysis, it was predicted that ORF2, ORF2a and ORF2ng are likely to be membrane proteins and so the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF2-1 (16kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 3A shows the results of affinity purification of the GST-fusion protein, and Figure 3B shows the results of expression of the His-fusion in *E.coli.* Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blots (Figure 3C), ELISA (positive result), and FACS analysis (Figure 3D). These experiments confirm that ORF37-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 10

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 77>:

```
  1   ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
 51   CGC.TGCGGG ACACTGACAG GTATTCCATC GCATGGCGgA GkTAAACgCT
101   TTgCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
151   GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
201   CACTATGGGC GACCAAGGTT CAGGcAGTTT GACAGGGGGG TCGCTACTCC
251   ATTGATGCAC kGrTwCsTGG CGAATACATA AACAGCCCTG CCGTCCGTAC
301   CGATTACACC TATCCACGTT ACGAAACCAC CGCTGAAACA ACATCAGGCG
351   GTTTGACAGG TTTAACCACT TCTTTATCTA CACTTAATGC CCCTGCACTC
401   TCTCGCACCC AATCAGACGG TAGCGGAAGT AAAAGCAGTC TGGGCTTAAA
451   TATTGGCGGG ATGGGGGATT ATCGAAATGA AACCTTGACG ACTAACCCGC
```

```
501  GCGACACTGC CTTTCTTTCC CACTTGGTAC AGACCGTATT TTTCCTGCGC
551  GGCATAGACG TTGTTTCTCC TGCCAATGCC GATACAGATG TGTTTATTAA
601  CATCGACGTA TTCGGAACGA TACGCAACAG AACCGAAATG..
```

This corresponds to the amino acid sequence <SEQ ID 78; ORF15>:

```
  1  MQARLLIPIL FSVFILSACG TLTGIPSHGG XKRFAVEQEL VAASARAAVK
 51  DMDLQALHGR KVALYIATMG DQGSGSLTGG RYSIDAXXXG EYINSPAVRT
101  DYTYPRYETT AETTSGGLTG LTTSLSTLNA PALSRTQSDG SGSKSSLGLN
151  IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
201  IDVFGTIRNR TEM..
```

Further work revealed the complete nucleotide sequence <SEQ ID 79>:

```
  1  ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
 51  CGCCTGCGGG ACACTGACAG GTATTCCATC GCATGGCGGA GGTAAACGCT
101  TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
151  GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
201  CACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
251  TTGATGCACT GATTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC
301  GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
351  TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
401  CTCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
451  ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
501  CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCCTGCGCG
551  GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAC
601  ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
651  TGCCGAAACA CTGAAAGCCC AAACAAAACT GGAATATTTC GCAGTAGACA
701  GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT
751  GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA
801  AGGAATTAAA CCGACGGAAG GATTAATGGT CGATTTCTCC GATATCCGAC
851  CATACGGCAA TCATACGGGT AACTCCGCCC CATCCGTAGA GGCTGATAAC
901  AGTCATGAGG GGTATGGATA CAGCGATGAA GTAGTGCGAC AACATAGACA
951  AGGACAACCT TGA
```

This corresponds to the amino acid sequence <SEQ ID 80; ORF15-1>:

```
  1  MQARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK
 51  DMDLQALHGR KVALYIATMG DQGSGSLTGG RYSIDALIRG EYINSPAVRT
101  DYTYPRYETT AETTSGGLTG LTTSLSTLNA PALSRTQSDG SGSKSSLGLN
151  IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
201  IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
251  AYKENYALWM GPYKVSKGIK PTEGLMVDFS DIRPYGNHTG NSAPSVEADN
301  SHEGYGYSDE VVRQHRQGQP *
```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 81>:

```
  1  ATGCAAGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
 51  CGCCTGCGGG ACACTGACAG GTATTCCATC GCATGGCGGA GGTAAACGCT
101  TTGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
151  GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
201  AACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
251  TTGATGCACT GATTCGTGGC GAATACATAA ACAGCCCTGC CGTCCGTACC
301  GATTACACCT ATCCACGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
351  TTTGACAGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
401  CGCGCACCCA ATCAGACGGT AGCGGAAGTA AAAGCAGTCT GGGCTTAAAT
451  ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CTAACCCGCG
501  CGACACTGCC TTTCTTTCCC ACTTGGTACA GACCGTATTT TTCCTGCGCG
551  GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACGGATGT GTTTATTAAC
601  ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
651  TGCCGAAACA CTGAAAGCCC AAACAAAACT GGAATATTTC GCAGTAGACA
701  GAACCAATAA AAAATTGCTC ATCAAACCAA AAACCAATGC GTTTGAAGCT
751  GCCTATAAAG AAAATTACGC ATTGTGGATG GGACCGTATA AAGTAAGCAA
801  AGGAATTAAA CCGACAGAAG GATTAATGGT CGATTTCTCC GATATCCAAC
851  CATACGGCAA TCATATGGGT AACTCTGCCC CATCCGTAGA GGCTGATAAC
901  AGTCATGAGG GGTATGGATA CAGCGATGAA GCAGTGCGAC GACATAGACA
951  AGGCAACCT TGA
```

This encodes a protein having amino acid sequence <SEQ ID 82; ORF15a>:

```
  1  MQARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK
```

```
 51  DMDLQALHGR KVALYIATMG DQGSGSLTGG RYSIDALIRG EYINSPAVRT
101  DYTYPRYETT AETTSGGLTG LTTSLSTLNA PALSRTQSDG SGSKSSLGLN
151  IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
201  IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
251  AYKENYALWM GPYKVSKGIK PTEGLMVDFS DIQPYGNHMG NSAPSVEADN
301  SHEGYGYSDE AVRRHRQGQP *
```

The originally-identified partial strain B sequence (ORF15) shows 98.1% identity over a 213aa overlap with ORF15a:

```
                   10        20        30        40        50        60
orf15.pep  MQARLLIPILFSVFILSACGTLTGIPSHGGXKRFAVEQELVAASARAAVKDMDLQALHGR
           |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
orf15a     MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAAVKDMDLQALHGR
                   10        20        30        40        50        60

                   70        80        90       100       110       120
orf15.pep  KVALYIATMGDQGSGSLTGGRYSIDAXXXGEYINSPAVRTDYTYPRYETTAETTSGGLTG
           ||||||||||||||||||||||||||   |||||||||||||||||||||||||||||||
orf15a     KVALYIATMGDQGSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGGLTG
                   70        80        90       100       110       120

                  130       140       150       160       170       180
orf15.pep  LTTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTTNPRDTAFLSHLVQTVF
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf15a     LTTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTTNPRDTAFLSHLVQTVF
                  130       140       150       160       170       180

                  190       200       210
orf15.pep  FLRGIDVVSPANADTDVFINIDVFGTIRNRTEM
           |||||||||||||||||||||||||||||||||
orf15a     FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
                  190       200       210       220       230       240
```

The complete strain B sequence (ORF15-1) and ORF15a show 98.8% identity in 320 aa overlap:

```
                     10        20        30        40        50        60
orf15a.pep   MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAAVKDMDLQALHGR
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf15-1      MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAAVKDMDLQALHGR
                     10        20        30        40        50        60

                     70        80        90       100       110       120
orf15a.pep   KVALYIATMGDQGSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGGLTG
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf15-1      KVALYIATMGDQGSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGGLTG
                     70        80        90       100       110       120

                    130       140       150       160       170       180
orf15a.pep   LTTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTTNPRDTAFLSHLVQTVF
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf15-1      LTTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTTNPRDTAFLSHLVQTVF
                    130       140       150       160       170       180

                    190       200       210       220       230       240
orf15a.pep   FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf15-1      FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
                    190       200       210       220       230       240

                    250       260       270       280       290       300
orf15a.pep   IKPKTNAFEAAYKENYALWMGPYKVSKGIKPTEGLMVDFSDIQPYGNHMGNSAPSVEADN
             ||||||||||||||||||||||||||||||||||||||||||:||||| |||||||||||
orf15-1      IKPKTNAFEAAYKENYALWMGPYKVSKGIKPTEGLMVDFSDIRPYGNHTGNSAPSVEADN
                    250       260       270       280       290       300

                    310       320
orf15a.pep   SHEGYGYSDEAVRRHRQGQPX
             ||||||||||:||:|||||||
orf15-1      SHEGYGYSDEVVRQHRQGQPX
```

```
            310       320
```

Further work identified the corresponding gene in *N.gonorrhoeae* <SEQ ID 83>:

```
  1 ATGCGGGCAC GGCTGCTGAT ACCTATTCTT TTTTCAGTTT TTATTTTATC
 51 CGCCTGCGGG ACACTGACAG GTATTCCATC GCATGGCGGA GGCAAACGCT
101 TCGCGGTCGA ACAAGAACTT GTGGCCGCTT CTGCCAGAGC TGCCGTTAAA
151 GACATGGATT TACAGGCATT ACACGGACGA AAAGTTGCAT TGTACATTGC
201 AACTATGGGC GACCAAGGTT CAGGCAGTTT GACAGGGGGT CGCTACTCCA
251 TTGATGCACT GATTCGCGGC GAATACATAA ACAGCCCTGC CGTCCGCACC
301 GATTACACCT ATCCGCGTTA CGAAACCACC GCTGAAACAA CATCAGGCGG
351 TTTGACGGGT TTAACCACTT CTTTATCTAC ACTTAATGCC CCTGCACTCT
401 CGCGCACCCA ATCAGACGGT AGCGGAAGTA GGAGCAGTCT GGGCTTAAAT
451 ATTGGCGGGA TGGGGGATTA TCGAAATGAA ACCTTGACGA CCAACCCGCG
501 CGACACTGCC TTTCTTTCCC ACTTGGTGCA GACCGTATTT TTCCTGCGCG
551 GCATAGACGT TGTTTCTCCT GCCAATGCCG ATACAGATGT GTTTATTAAC
601 ATCGACGTAT TCGGAACGAT ACGCAACAGA ACCGAAATGC ACCTATACAA
651 TGCCGAAACA CTGAAAGCCC AAACAAAACT GGAATATTTC GCAGTAGACA
701 GAACCAATAA AAAATTGCTC ATCAAACCCA AAACCAATGC GTTTGAAGCT
751 GCCTATAAAG AAAATTACGC ATTGTGGATG GGGCCGTATA AAGTAAGCAA
801 AGGAATCAAA CCGACGGAAG GATTGATGGT CGATTTCTCC GATATCCAAC
851 CATACGGCAA TCATACGGGT AACTCCGCCC CATCCGTAGA GGCTGATAAC
901 AGTCATGAGG GGTATGGATA CAGCGATGAA GCAGTGCGAC AACATAGACA
951 AGGGCAACCT TGA
```

This encodes a protein having amino acid sequence <SEQ ID 84; ORF15ng>:

```
  1 MRARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK
 51 DMDLQALHGR KVALYIATMG DQGSGSLTGG RYSIDALIRG EYINSPAVRT
101 DYTYPRYETT AETTSGGLTG LTTSLSTLNA PALSRTQSDG SGSRSSLGLN
151 IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF FLRGIDVVSP ANADTDVFIN
201 IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL IKPKTNAFEA
251 AYKENYALWM GPYKVSKGIK PTEGLMVDFS DIQPYGNHTG NSAPSVEADN
301 SHEGYGYSDE AVRQHRQGQP *
```

The originally-identified partial strain B sequence (ORF15) shows 97.2% identity over a 213aa overlap with ORF15ng:

```
orf15.pep  MQARLLIPILFSVFILSACGTLTGIPSHGGXKRFAVEQELVAASARAAVKDMDLQALHGR   60
           |:|||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
orf15ng    MRARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAAVKDMDLQALHGR   60

orf15.pep  KVALYIATMGDQGSGSLTGGRYSIDAXXXGEYINSPAVRTDYTYPRYETTAETTSGGLTG  120
           ||||||||||||||||||||||||||   |||||||||||||||||||||||||||||||
orf15ng    KVALYIATMGDQGSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGGLTG  120

orf15.pep  LTTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTTNPRDTAFLSHLVQTVF  180
           |||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||
orf15ng    LTTSLSTLNAPALSRTQSDGSGSRSSLGLNIGGMGDYRNETLTTNPRDTAFLSHLVQTVF  180

orf15.pep  FLRGIDVVSPANADTDVFINIDVFGTIRNRTEM                             213
           |||||||||||||||||||||||||||||||||
orf15ng    FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL  240
```

The complete strain B sequence (ORF15-1) and ORF15ng show 98.8% identity in 320 aa overlap:

```
                      10        20        30        40        50        60
orf15-1.pep  MQARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAAVKDMDLQALHGR
             |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf15ng      MRARLLIPILFSVFILSACGTLTGIPSHGGGKRFAVEQELVAASARAAVKDMDLQALHGR
                      10        20        30        40        50        60

                      70        80        90       100       110       120
orf15-1.pep  KVALYIATMGDQGSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGGLTG
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf15ng      KVALYIATMGDQGSGSLTGGRYSIDALIRGEYINSPAVRTDYTYPRYETTAETTSGGLTG
                      70        80        90       100       110       120

                     130       140       150       160       170       180
orf15-1.pep  LTTSLSTLNAPALSRTQSDGSGSKSSLGLNIGGMGDYRNETLTTNPRDTAFLSHLVQTVF
```

```
              ||||||||||||||||||||:|||||||||||||||||||||||||||||||||||
orf15ng       LTTSLSTLNAPALSRTQSDGSGSRSSLGLNIGGMGDYRNETLTTNPRDTAFLSHLVQTVF
                    130       140       150       160       170       180

                    190       200       210       220       230       240
orf15-1.pep   FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf15ng       FLRGIDVVSPANADTDVFINIDVFGTIRNRTEMHLYNAETLKAQTKLEYFAVDRTNKKLL
                    190       200       210       220       230       240

                    250       260       270       280       290       300
orf15-1.pep   IKPKTNAFEAAYKENYALWMGPYKVSKGIKPTEGLMVDFSDIRPYGNHTGNSAPSVEADN
              ||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||
orf15ng       IKPKTNAFEAAYKENYALWMGPYKVSKGIKPTEGLMVDFSDIQPYGNHTGNSAPSVEADN
                    250       260       270       280       290       300

                    310       320
orf15-1.pep   SHEGYGYSDEVVRQHRQGQPX
              ||||||||||:||||||||||
orf15ng       SHEGYGYSDEAVRQHRQGQPX
                    310       320
```

Computer analysis of these amino acid sequences reveals an ILSAC motif (putative membrane lipoprotein lipid attachment site, as predicted by the MOTIFS program).

indicates a putative leader sequence, and it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF15-1 (31.7kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 4A shows the results of affinity purification of the GST-fusion protein, and Figure 4B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for Western blot (Figure 4C) and ELISA (positive result). These experiments confirm that ORFX-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 11

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 85>:

```
  1   ..GG.CAGCACA AAAAACAGGC GGTTGAACGG AAAAACCGTA TTTACGATGA
 51     TGCCGGGTAT GATATTCGGC GTATTCACGG GCGCATTCTC CGCAAAATAT
101     ATCCCCGCGT TCGGGCTTCA AATTTTCTTC ATCCTGTTTT TAACCGCCGT
151     CGCATTCAAA ACACTGCATA CCGACCCTCA GACGGCATCC CGCCCGCTGC
201     CCGGACTGCC CrGACTGACT GCGGTTTCCA CACTGTTCGG CACAATGTCG
251     AGCTGGGTCG GCATAGGCGG CGGTTCACTT TCCGTCCCCT TCTTAATCCA
301     CTGCGGCTTC CCCGCCCATA AAGCCATCGG CACATCATCC GGCCTTGCCT
351     GGCCGATTGC ACTCTCCGGC GCAATATCGT ATCTGCTCAA CGGCCTGAAT
401     ATTGCAGGAT TGCCCGAAGG GTCACTGGGC TTCCTTTACC TGCCCGCCGT
451     CGCCGTCCTC AGCGCGGCAA CCATTGCCTT TGCCCCGCTC GGTGTCAAAA
501     CCGCCCACAA ACTTTCTTCT GCCAAACTCA AAAAATC.TT CGGCATTATG
551     TTGCTTTTGA TTGCCGGAAA AATGCTGTAC AACCTGCTTT AA
```

This corresponds to the amino acid sequence <SEQ ID 86; ORF17>:

```
  1   ..GQHKKQAVNG KTVFTMMPGM IFGVFTGAFS AKYIPAFGLQ IFFILFLTAV
 51     AFKTLHTDPQ TASRPLPGLP XLTAVSTLFG TMSSWVGIGG GSLSVPFLIH
```

```
101    CGFPAHKAIG TSSGLAWPIA LSGAISYLLN GLNIAGLPEG SLGFLYLPAV
151    AVLSAATIAF APLGVKTAHK LSSAKLKKSF GIMLLLIAGK MLYNLL*
```

Further work revealed the complete nucleotide sequence <SEQ ID 87>:

```
  1  ATGTGGCATT GGGACATTAT CTTAATCCTG CTTGCCGTAG GCAGTGCGGC
 51  AGGTTTTATT GCCGGCCTGT TCGGCGTAGG CGGCGGCACG CTGATTGTCC
101  CTGTCGTTTT ATGGGTGCTT GATTTGCAGG GTTTGGCACA ACATCCTTAC
151  GCGCAACACC TCGCCGTCGG CACATCCTTC GCCGTCATGG TCTTCACCGC
201  CTTTTCCAGT ATGCTGGGGC AGCACAAAAA ACAGGCGGTC GACTGGAAAA
251  CCGTATTTAC GATGATGCCG GGTATGATAT TCGGCGTATT CACGGGCGCA
301  CTCTCCGCAA AATATATCCC CGCGTTCGGG CTTCAAATTT TCTTCATCCT
351  GTTTTTAACC GCCGTCGCAT TCAAAACACT GCATACCGAC CCTCAGACGG
401  CATCCCGCCC GCTGCCCGGA CTGCCCGGAC TGACTGCGGT TTCCACACTG
451  TTCGGCACAA TGTCGAGCTG GGTCGGCATA GGCGGCGGTT CACTTTCCGT
501  CCCCTTCTTA ATCCACTGCG GCTTCCCCGC CCATAAAGCC ATCGGCACAT
551  CATCCGGCCT TGCCTGGCCG ATTGCACTCT CCGGCGCAAT ATCGTATCTG
601  CTCAACGGCC TGAATATTGC AGGATTGCCC GAAGGGTCAC TGGGCTTCCT
651  TTACCTGCCC GCCGTCGCCG TCCTCAGCGC GGCAACCATT GCCTTTGCCC
701  CGCTCGGTGT CAAAACCGCC CACAAACTTT CTTCTGCCAA ACTCAAAAAA
751  Tc.TTCGGCA TTATGTTGCT TTTGATTGCC GGAAAAATGC TGTACAACCT
801  GCTTTAA
```

This corresponds to the amino acid sequence <SEQ ID 88; ORF17-1>:

```
  1  MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY
 51  AQHLAVGTSF AVMVFTAFSS MLGQHKKQAV DWKTVFTMMP GMIFGVFTGA
101  LSAKYIPAFG LQIFFILFLT AVAFKTLHTD PQTASRPLPG LPGLTAVSTL
151  FGTMSSWVGI GGGSLSVPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL
201  LNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKK
251  XFGIMLLLIA GKMLYNLL*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with hypothetical *H.influenzae* transmembrane protein HI0902 (accession number P44070)

ORF17 and HI0902 proteins show 28% aa identity in 192 aa overlap:

```
ORF17     3   HKKQAVNGKTVFTMMPGMIFGVFT-GAFSAKYIPAFGLQIF--FILFLTAVAFKTLHTDP 59
              HK   +  + V  + P ++  VF  G F  +       +IF   +++L      ++  D
HI0902   72   HKLGNIVWQAVRILAPVIMLSVFICGLFIGRLDREISAKIFACLVVYLATKMVLSIKKD- 130

ORF17    60   QTASRPLPGLPXLTAVSTLFGTMSSWVGIGGGSLSVPFLIHCGFPAHKAIGTSSGLAWPI 119
              Q  ++ L  L  +      L G  SS  GIGGG   VPFL   G     +AIG+S+      +
HI0902  131   QVTTKSLTPLSSVIG-GILIGMASSAAGIGGGGFIVPFLTARGINIKQAIGSSAFCGMLL 189

ORF17   120   ALSGAISYLLNGLNIAGLPEGSLGFLYLPAVAVLSAATIAFAPLGVXXXXXXXXXXXXXX 179
               +SG  S++++G      +PE SLG++YLPAV  ++A +    + LG
HI0902  190   GISGMFSFIVSGWGNPLMPEYSLGYIYLPAVLGITATSFFTSKLGASATAKLPVSTLKKG 249

ORF17   180   FGIMLLLIAGKM 191
              F + L+++A  M
HI0902  250   FALFLIVVAINM 261
```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF17 shows 96.9% identity over a 196aa overlap with an ORF (ORF17a) from strain A of *N. meningitidis:*

```
                                                  10        20        30
orf17.pep                                 GQHKKQAVNGKTVFTMMPGMIFGVFTGAFS
                                          ||||||||: ||||||||||:||||:||:|
orf17a      QGLAQHPYAQHLAVGTSFAVMVFTAFSSMLGQHKKQAVDWKTVFTMMPGMVFGVFAGALS
                    50        60        70        80        90       100

                    40        50        60        70        80        90
orf17.pep   AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPXLTAVSTLFGTMSSWVGIGG
            |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
orf17a      AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSWVGIGG
```

```
                  110       120       130       140       150       160

                    100       110       120       130       140       150
orf17.pep   GSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPAV
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf17a      GSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPAV
                  170       180       190       200       210       220

                    160       170       180       190
orf17.pep   AVLSAATIAFAPLGVKTAHKLSSAKLKKSFGIMLLLIAGKMLYNLLX
            |||||||||||||||||||||||||||||||||||||||||||||||
orf17a      AVLSAATIAFAPLGVKTAHKLSSAKLKKSFGIMLLLIAGKMLYNLLX
                  230       240       250       260
```

The complete length ORF17a nucleotide sequence <SEQ ID 89> is:

```
  1 ATGTGGCATT GGGACATTAT CTTAATCCTG CTTGCCGTAG GCAGTGCGGC
 51 AGGTTTTATT GCCGGCCTGT TCGGCGTAGG CGGCGGCACG CTGATTGTCC
101 CTGTCGTTTT ATGGGTGCTT GATTTGCAGG GTTTGGCACA ACATCCTTAC
151 GCGCAACACC TCGCCGTCGG CACATCCTTC GCCGTCATGG TCTTCACCGC
201 CTTTTCCAGT ATGCTGGGGC AGCACAAAAA ACAGGCGGTC GACTGGAAAA
251 CCGTATTTAC GATGATGCCG GGTATGGTAT TCGGCGTATT CGCTGGCGCA
301 CTCTCCGCAA AATATATCCC AGCGTTCGGG CTTCAAATTT TCTTCATCCT
351 GTTTTTAACC GCCGTCGCAT TCAAAACACT GCATACCGAC CCTCAGACGG
401 CATCCCGCCC GCTGCCCGGA CTGCCCGGAC TGACTGCGGT TTCCACACTG
451 TTCGGCACAA TGTCGAGCTG GGTCGGCATA GGCGGCGGTT CACTTTCCGT
501 CCCCTTCTTA ATCCACTGCG GCTTCCCCGC CCATAAAGCC ATCGGCACAT
551 CATCCGGCCT TGCCTGGCCG ATTGCACTCT CCGGCGCAAT ATCGTATCTG
601 CTCAACGGCC TGAATATTGC AGGATTGCCC GAAGGGTCAC TGGGCTTCCT
651 TTACCTGCCC GCCGTCGCCG TCCTCAGCGC GGCAACCATT GCCTTTGCCC
701 CGCTCGGTGT CAAAACCGCC CACAAACTTT CTTCTGCCAA ACTCAAAAAA
751 TCCTTCGGCA TTATGTTGCT TTTGATTGCC GGAAAAATGC TGTACAACCT
801 GCTTTAA
```

This encodes a protein having amino acid sequence <SEQ ID 90>:

```
  1 MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY
 51 AQHLAVGTSF AVMVFTAFSS MLGQHKKQAV DWKTVFTMMP GMVFGVFAGA
101 LSAKYIPAFG LQIFFILFLT AVAFKTLHTD PQTASRPLPG LPGLTAVSTL
151 FGTMSSWVGI GGGSLSVPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL
201 LNGLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKK
251 SFGIMLLLIA GKMLYNLL*
```

ORF17a and ORF17-1 show 98.9% identity in 268 aa overlap:

```
                      10        20        30        40        50        60
orf17a.pep    MWHWDIILILLAVGSAAGFIAGLFGVGGGTLIVPVVLWVLDLQGLAQHPYAQHLAVGTSF
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf17-1       MWHWDIILILLAVGSAAGFIAGLFGVGGGTLIVPVVLWVLDLQGLAQHPYAQHLAVGTSF
                      10        20        30        40        50        60

                      70        80        90       100       110       120
orf17a.pep    AVMVFTAFSSMLGQHKKQAVDWKTVFTMMPGMVFGVFAGALSAKYIPAFGLQIFFILFLT
              |||||||||||||||||||||||||||||||||:||||:|||||||||||||||||||||
orf17-1       AVMVFTAFSSMLGQHKKQAVDWKTVFTMMPGMIFGVFTGALSAKYIPAFGLQIFFILFLT
                      70        80        90       100       110       120

                     130       140       150       160       170       180
orf17a.pep    AVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSWVGIGGGSLSVPFLIHCGFPAHKA
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf17-1       AVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSWVGIGGGSLSVPFLIHCGFPAHKA
                     130       140       150       160       170       180

                     190       200       210       220       230       240
orf17a.pep    IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPAVAVLSAATIAFAPLGVKTA
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf17-1       IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPAVAVLSAATIAFAPLGVKTA
                     190       200       210       220       230       240

                     250       260      269
orf17a.pep    HKLSSAKLKKSFGIMLLLIAGKMLYNLLX
```

```
                ||||||||| |||||||||||||||||
orf17-1         HKLSSAKLKKXFGIMLLLIAGKMLYNLLX
                        250       260
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF17 shows 93.9% identity over a 196aa overlap with a predicted ORF (ORF17.ng) from *N. gonorrhoeae*:

```
orf17.pep                                          GQHKKQAVNGKTVFTMMPGMIFGVFTCAFS   30
                                                   |||||||: ||:|:||||||||||:||:|
orf17ng     QGLAQHPYAQHLAVGTSFAVMVFTAFSSMLGQHKKQAVDWKTIFAMMPGMIFGVFAGALS  102

orf17.pep   AKYIPAFGLQIFFILFLTAVAFKTLHTDPQTASRPLPGLPXLTAVSTLFGTMSSWVCICC   90
            ||||||||||||||||||||||||||  ||||||||||| |||||||||:||||||||
orf17ng     AKYIPAFGLQIFFILFLTAVAFKTLHTGRQTASRPLPGLPGLTAVSTLFGAMSSWVCICC  162

orf17.pep   GSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPAV  150
            ||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
orf17ng     GSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLPAV  202

orf17.pep   AVLSAATIAFAPLGVKTAHKLSSAKLKKSFGIMLLLIAGKMLYNLL  196
            ||||||||||||||||||||||||||:||||||||||||||||||
orf17ng     AVLSAATIAFAPLGVKTAHKLSSAKLKESFGIMLLLIAGKMLYNLL  268
```

An ORF17ng nucleotide sequence <SEQ ID 91> is predicted to encode a protein having amino acid sequence <SEQ ID 92>:

```
  1  MWHWDIILIL LAVGSAAGFI AGLFGVGGGT LIVPVVLWVL DLQGLAQHPY
 51  AQHLAVGTSF AVMVFTAFSS MLGQHKKQAV DWKTIFAMMP GMIFGVFAGA
101  LSAKYIPAFG LQIFFILFLT AVAFKTLHTG RQTASRPLPG LPGLTAVSTL
151  FGAMSSWVGI GGGSLSVPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL
201  VNCLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKE
251  SFCIMLLLIA GKMLYNLL*
```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 93>:

```
  1  ATGTGGCATT GGGACATTAT CTTAATCCTG CTTGCcgtag gcAGTGCCGC
 51  AGGTTTTATT GCCGGCCTGT Tcggtgtagg cggcgGTACG CTGATTGTCC
101  CTGTCGTTTT ATGGGTGCTT GATTTGCAGG GTTTGGCACA ACATCCTTAC
151  GCGCAACACC TCGCCGTCGG CAcaTccttc gcCGTCATGG TCTTCACCGC
201  CTTTTCCAGT ATGTTGGGGC AGCACAAAAA ACAGGCGGTC GACTGGAAAA
251  CCATATTTGC GATGATGCCG GGTATGATAT TCGGCGTATT CGCTGGCGCA
301  CTCTCCGCAA AATATATCCC CGCGTTCGGG CTTCAAATTT TCTTCATCCT
351  GTTTTTAACC GCCGTCGCAT TCAAAACACT GCATACCGGT CGTCAGACGG
401  CATCCCGCCC GCTGCCCGGG CTGCCCGGAC TGACTGCGGT TTCCACACTG
451  TTCGGCGCAA TGTCGAGCTG GGTCGGCATA GGCGGCGGTT CACTTTCCGT
501  CCCCTTCTTA ATCCACTGCG GCTTCCCCGC CCATAAAGCC ATCGGCACAT
551  CATCCGGCCT TGCCTGGCCG ATTGCACTCT CCGGCGCAAT ATCGTATCTG
601  GTCAACGGTC TGAATATTGC AGGATTGCCC GAAGGGTCGC TGGGCTTCCT
651  TTACCTGCCC GCCGTCGCCG TCCTCAGCGC GGCAACCATT GCCTTTGCCC
701  CCCTCGGTGT CAAAACCGCC CACAAACTTT CTTCTGCCAA ACTCAAAGAA
751  TCCTTCGGCA TTATGTTGCT TTTGATTGCC GGAAAAATGC TGTACAACCT
801  GCTTTAA
```

This corresponds to the amino acid sequence <SEQ ID 94; ORF17ng-1>:

```
  1  MWHWDIILIL LAVCSAAGFI AGLFCVCGGT LIVPVVLWVL DLQGLAQHPY
 51  AQHLAVGTSF AVMVFTAFSS MLCQHKKQAV DWKTIFAMMP GMIFGVFAGA
101  LSAKYIPAFC LQIFFILFLT AVAFKTLHTG RQTASRPLPG LPGLTAVSTL
151  FCAMSSWVGI CGGSLSVPFL IHCGFPAHKA IGTSSGLAWP IALSGAISYL
201  VNCLNIAGLP EGSLGFLYLP AVAVLSAATI AFAPLGVKTA HKLSSAKLKE
251  SFGIMLLLIA CKMLYNLL*
```

ORF17ng-1 and ORF17-1 show 96.6% identity in 268 aa overlap:

```
                     10        20        30        40        50        60
orf17-1.pep  MWHWDIILILLAVGSAAGFIAGLFGVGGGTLIVPVVLWVLDLQGLAQHPYAQHLAVGTSF
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf17ng-1     MWHWDIIILILLAVGSAAGFIAGLFGVGGGTLIVPVVLWVLDLQGLAQHPYAQHLAVGTSF
                      10        20        30        40        50        60

                      70        80        90       100       110       120
orf17-1.pep   AVMVFTAFSSMLGQHKKQAVDWKTVFTMMPGMIFGVFTGALSAKYIPAFGLQIFFILFLT
              ||||||||||||||||||||||||:|:||||||||||:||||||||||||||||||||||
orf17ng-1     AVMVFTAFSSMLGQHKKQAVDWKTIFAMMPGMIFGVFAGALSAKYIPAFGLQIFFILFLT
                      70        80        90       100       110       120

                     130       140       150       160       170       180
orf17-1.pep   AVAFKTLHTDPQTASRPLPGLPGLTAVSTLFGTMSSWVGIGGGSLSVPFLIHCGFPAHKA
              |||||||||  |||||||||||||||||||||:||||||||||||||||||||||||||||
orf17ng-1     AVAFKTLHTGRQTASRPLPGLPGLTAVSTLFGAMSSWVGIGGGSLSVPFLIHCGFPAHKA
                     130       140       150       160       170       180

                     190       200       210       220       230       240
orf17-1.pep   IGTSSGLAWPIALSGAISYLLNGLNIAGLPEGSLGFLYLPAVAVLSAATIAFAPLGVKTA
              ||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
orf17ng-1     IGTSSGLAWPIALSGAISYLVNGLNIAGLPEGSLGFLYLPAVAVLSAATIAFAPLGVKTA
                     190       200       210       220       230       240

                     250       260      269
orf17-1.pep   HKLSSAKLKKXFGIMLLLIAGKMLYNLLX
              ||||||||||: |||||||||||||||||
orf17ng-1     HKLSSAKLKESFGIMLLLIAGKMLYNLLX
                     250       260
```

In addition, ORF17ng-1 shows significant homology with a hypothetical *H.influenzae* protein:

```
sp|P44070|Y902_HAEIN HYPOTHETICAL PROTEIN HI0902 pir||G64015 hypothetical protein
HI0902 - Haemophilus influenzae (strain Rd KW20) gi|1573922 (U32772) H. influenzae
predicted coding region HI0902 [Haemophilus influenzae]Length = 264
 Score = 74 (34.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23
 Identities = 15/43 (34%), Positives = 23/43 (53%)

Query:    55 AVGTSFAVMVFTAFSSMLGQHKKQAVDWKTIFAMMPGMIFGVF 97
             A+GTSFA +V T   S    HK    + W+ +   + P ++  VF
Sbjct:    52 ALGTSFATIVITGIGSAQRHHKLGNIVWQAVRILAPVIMLSVF 94

 Score = 195 (91.9 bits), Expect = 1.6e-23, Sum P(2) = 1.6e-23
 Identities = 44/114 (38%), Positives = 65/114 (57%)

Query:   150 LFGAMSSWVGIGGGSLSVPFLIHCGFPAHKAIGTSSGLAWPIALSGAISYLVNGLNIAGL 209
             L G  SS  GIGGG    VPFL   G     +AIG+S+      + +SG  S++V+G      +
Sbjct:   148 LIGMASSAAGIGGGGFIVPFLTARGINIKQAIGSSAFCGMLLGISGMFSFIVSGWGNPLM 207

Query:   210 PEGSLGFLYLPAVAVLSAATIAFAPLGVKTAHKLSSAKLKESFGIMLLLIAGKM 263
             PE SLG++YLPAV  ++A +    + LG      KL  + LK+ F + L++A  M
Sbjct:   208 PEYSLGYIYLPAVLGITATSFFTSKLGASATAKLPVSTLKKGFALFLIVVAINM 261
```

This analysis, including the homology with the hypothetical *H.influenzae* transmembrane protein, suggests that the proteins from *N.meningitidis* and *N.gonorrhoeae,* and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 12

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 95>:

```
  1  ..GGAAACGGAT GGCAGGCAGA CCCCGAACAT CCGCTGCTCG GGCTTTTTGC
 51    CGTCAGTAAT GTATCGATGA CGCTTGCTTT TGTCGGAATA TGTGCGTTGG
101    TGCATTATTG CTTTTCGGGA ACGGTTCAAG TGTTTGTGTT TGCGGCACTG
151    CTCAAACTTT ATGCGCTGAA GCCGGTTTAT TGGTTCGTGT TGCAGTTTGT
201    GCTGATGGCG GTTGCCTATG TCCACCGCTG CGGTATAGAC CGGCAGCCGC
251    CGTCAACGTT CGGCGGCTCG CAGCTGCGAC TCGGCGGGTT GACGGCAGCG
```

```
301    TTGATGCAGG TCTCGGTACT GGTGCTGCTG CTTTCAGAAA TTGGAAGATA
351    A
```

This corresponds to the amino acid sequence <SEQ ID 96; ORF18>:

```
  1  ..GNGWQADPEH PLLGLFAVSN VSMTLAFVGI CALVHYCFSG TVQVFVFAAL
 51    LKLYALKPVY WFVLQFVLMA VAYVHRCGID RQPPSTFGGS QLRLGGLTAA
101    LMQVSVLVLL LSEIGR*
```

Further work revealed the complete nucleotide sequence <SEQ ID 97>:

```
  1  ATGATTTTGC TGCATTTGGA TTTTTTGTCT GCCTTACTGT ATGCGGCGGT
 51  TTTTCTGTTT CTGATATTCC GCGCAGGAAT GTTGCAATGG TTTTGGGCGA
101  GTATTATGCT GTGGCTGGGC ATATCGGTTT TGGGGGCAAA GCTGATGCCC
151  GGCATATGGG GAATGACCCG CGCCGCGCCC TTGTTCATCC CCCATTTTTA
201  CCTGACTTTG GGCAGCATAT TTTTTTTCAT CGGGCATTGG AACCGGAAAA
251  CAGATGGAAA CGGATGGCAG GCAGACCCCG AACATCCGCT GCTCGGGCTT
301  TTTGCCGTCA GTAATGTATC GATGACGCTT GCTTTTGTCG GAATATGTGC
351  GTTGGTGCAT TATTGCTTTT CGGGAACGGT TCAAGTGTTT GTGTTTGCGG
401  CACTGCTCAA ACTTTATGCG CTGAAGCCGG TTTATTGGTT CGTGTTGCAG
451  TTTGTGCTGA TGGCGGTTGC CTATGTCCAC CGCTGCGGTA TAGACCGGCA
501  GCCGCCGTCA ACGTTCGGCG GCTCGCAGCT GCGACTCGGC GGGTTGACGG
551  CAGCGTTGAT GCAGGTCTCG GTACTGGTGC TGCTGCTTTC AGAAATTGGA
601  AGATAA
```

This corresponds to the amino acid sequence <SEQ ID 98; ORF18-1>:

```
  1  MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIMLWLG ISVLGAKLMP
 51  GIWGMTRAAP LFIPHFYLTL GSIFFFIGHW NRKTDGNGWQ ADPEHPLLGL
101  FAVSNVSMTL AFVGICALVH YCFSGTVQVF VFAALLKLYA LKPVYWFVLQ
151  FVLMAVAYVH RCGIDRQPPS TFGGSQLRLG GLTAALMQVS VLVLLLSEIG
201  R*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF18 shows 98.3% identity over a 116aa overlap with an ORF (ORF18a) from strain A of *N. meningitidis:*

```
                                                   10        20        30
orf18.pep                              GNGWQADPEHPLLGLFAVSNVSMTLAFVGI
                                       ||||||||||||||||||||||||||||||
orf18a      TRAAPLFIPHFYLTLGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGI
              60        70        80        90       100       110

               40        50        60        70        80        90
orf18.pep   CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGS
            ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
orf18a      CALVHYCFSXTVQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGS
              120       130       140       150       160       170

                 100       110
orf18.pep   QLRLGGLTAALMQVSVLVLLLSEIGRX
            ||||||||||||| ||||||||||||
orf18a      QLRLGGLTAALMQXSVLVLLLSEIGRX
              180       190       200
```

The complete length ORF18a nucleotide sequence <SEQ ID 99> is:

```
  1  ATGATTTTGC TGCATTTGGA TTTTTTGTCT GCCTTACTGT ATGCGGCGGT
 51  TTTTCTGTTT CTGATATTCC GCGCAGGAAT GTTGCAATGG TTTTGGGCGA
101  GTATTATGCT GTGGCTGGGC ATATCGGTTT TGGGGGCAAA GCTGATGCCC
151  GGCATATGGG GAATGACCCG CGCCGCGCCC TTGTTCATCC CCCATTTTTA
201  CCTGACTTTG GGCAGCATAT TTTTTTTCAT CGGGCATTGG AACCGGAAAA
251  CGGATGGAAA CGGATGGCAG GCAGACCCCG AACATCCTCT GCTCGGGCTG
301  TTTGCCGTCA GTAATGTATC GATGACGCTT GCTTTTGTCG GAATATGTGC
351  GTTGGTGCAT TATTGCTTTT CGNGAACGGT TCAAGTGTTT GTGTTTGCGG
401  CACTGCTCAA ACTTTATGCG CTGAAGCCGG TTTATTGGTT CGTGTTGCAG
```

```
451  TTTGTGCTGA TGGCGGTTGC CTATGTCCAC CGCTGCGGTA TAGACCGGCA
501  GCCGCCGTCA ACGTTCGGCG GNTCGCAGCT GCGACTCGGC GGGTTGACGG
551  CAGCGTTGAT GCAGNTCTCG GTACTGGTGC TGCTGCTTTC AGAAATTGGA
601  AGATAA
```

This encodes a protein having amino acid sequence <SEQ ID 100>:

```
  1  MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIMLWLG ISVLGAKLMP
 51  GIWGMTRAAP LFIPHFYLTL GSIFFFIGHW NRKTDGNGWQ ADPEHPLLGL
101  FAVSNVSMTL AFVGICALVH YCFSXTVQVF VFAALLKLYA LKPVYWFVLQ
151  FVLMAVAYVH RCGIDRQPPS TFGGSQLRLG GLTAALMQXS VLVLLLSEIG
201  R*
```

ORF18a and ORF18-1 show 99.0% identity in 201 aa overlap:

```
                    10        20        30        40        50        60
orf18a.pep   MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf18-1      MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP
                    10        20        30        40        50        60

                    70        80        90       100       110       120
orf18a.pep   LFIPHFYLTLGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGICALVH
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf18-1      LFIPHFYLTLGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGICALVH
                    70        80        90       100       110       120

                   130       140       150       160       170       180
orf18a.pep   YCFSXTVQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGSQLRLG
             |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf18-1      YCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGSQLRLG
                   130       140       150       160       170       180

                   190       200
orf18a.pep   GLTAALMQXSVLVLLLSEIGRX
             |||||||| ||||||||||||
orf18-1      GLTAALMQVSVLVLLLSEIGRX
                   190       200
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF18 shows 93.1% identity over a 116aa overlap with a predicted ORF (ORF18.ng) from *N. gonorrhoeae:*

```
orf18.pep                                 GNGWQADPEHPLLGLFAVSNVSMTLAFVGI  30
                                          ||||||||||||||||||||||||||||||
orf18ng     TRAAPLFIPHFYLTLGSIFFFIGYWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGI  115

orf18.pep   CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGS   90
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf18ng     CALVHYCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGS  175

orf18.pep   QLRLGGLTAALMQVSVLVLLLSEIGR  116
            ||||| |:| ||||:| ::||:||||
orf18ng     QLRLGVLAAMLMQVAVTAMLLAEIGR  201
```

The complete length ORF18ng nucleotide sequence is <SEQ ID 101>:

```
  1  ATGATTTTGC TGCATTTGGA TTTTTTGTCT GCCTTACTGt aTGCGGcggt
 51  tttTctgTTT CTGATATTCC GCGCAGGAAT GTTGCAATGG TTTTGGGCGA
101  GTATTGCGTT GTGGCTCGGC ATCTCGGTTT TAGGGGTAAA GCTGATGCCG
151  GGGATGTGGG GAATGACCCG CGCCGCGCCT TTGTTCATCC CCCATTTTTA
201  CCTGACTTTG GGCAGCATAT TTTTTTTCAT CGGGTATTGG AACCGGAAAA
251  CAGATGGAAA CGGATGGCAG GCAGACCCCG AACATCCGCT GCTCGGGCTT
301  TTTGCCGTCA GTAATGTATC GATGACGCTT GCTTTTGTCG GAATATGTGC
351  GTTGGTGCAT TATTGCTTTT CGGGAACGGT TCAAGTGTTT GTGTTTGCGG
401  CATTGCTCAA ACTTTATGCG CTGAAGCCGG TTTATTGGTT CGTGTTGCAG
451  TTTGTATTGA TGGCGGttgC CTATGTCCAC CGCTGCGGTA TAGACCGGCA
501  GCCGCCGTCA ACGTTCGGCG GTTCGCAGCT GCGACTCGGC GTGTTGGCGG
```

```
551  CGATGTTGAT GCAGGTTGCG GTAACGGCGA TGCTGCTTGC CGAAATCGGC
601  AGATGA
```

This encodes a protein having amino acid sequence <SEQ ID 102>:

```
  1  MILLHLDFLS ALLYAAVFLF LIFRAGMLQW FWASIALWLG ISVLGVKLMP
 51  GMWGMTRAAP LFIPHFYLTL GSIFFFIGYW NRKTDGNGWQ ADPEHPLLGL
101  FAVSNVSMTL AFVGICALVH YCFSGTVQVF VFAALLKLYA LKPVYWFVLQ
151  FVLMAVAYVH RCGIDRQPPS TFGGSQLRLG VLAAMLMQVA VTAMLLAEIG
201  R*
```

This ORF18ng protein sequence shows 94.0% identity in 201 aa overlap with ORF18-1:

```
                       10        20        30        40        50        60
orf18-1.pep    MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIMLWLGISVLGAKLMPGIWGMTRAAP
               ||||||||||||||||||||||||||||||||||| |||||||||:|||||:||||||||
orf18ng        MILLHLDFLSALLYAAVFLFLIFRAGMLQWFWASIALWLGISVLGVKLMPGMWGMTRAAP
                       10        20        30        40        50        60

                       70        80        90       100       110       120
orf18-1.pep    LFIPHFYLTLGSIFFFIGHWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGICALVH
               ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
orf18ng        LFIPHFYLTLGSIFFFIGYWNRKTDGNGWQADPEHPLLGLFAVSNVSMTLAFVGICALVH
                       70        80        90       100       110       120

                      130       140       150       160       170       180
orf18-1.pep    YCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGSQLRLG
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf18ng        YCFSGTVQVFVFAALLKLYALKPVYWFVLQFVLMAVAYVHRCGIDRQPPSTFGGSQLRLG
                      130       140       150       160       170       180

                      190       200
orf18-1.pep    GLTAALMQVSVLVLLLSEIGRX
                |:| ||||:| ::||:|||||
orf18ng        VLAAMLMQVAVTAMLLAEIGRX
                      190       200
```

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 13

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 103>:

```
  1  ATGAAAACCC CACTCCTCAA GCCTCTGCTN ATTACCTCGC TTCCCGTTTT
 51  CGCCAGTGTT TTTACCGCCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA
101  AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CCTTGTCGAT
151  TTGGACAACC NCNTGACCGG ACGGCTNAAA AACATCATCA CCACCGTCGC
201  CCTGTTCACC CTCTCCTCGC TCACGGCACA AAGCACCCTC GGCACAGGGC
251  TGCCCTTCAT CCTCGCCATG ACCCTGATGA CTT.CG.CTT CACCATTTTA
301  GGCGCGGNCG ...
```

This corresponds to the amino acid sequence <SEQ ID 104; ORF19>:

```
  1  MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
 51  LDNXXTGRLK NIITTVALFT LSSLTAQSTL GTGLPFILAM TLMTXXFTIL
101  GAX...
```

Further work revealed the complete nucleotide sequence <SEQ ID 105>:

```
  1  ATGAAAACCC CACTCCTCAA GCCTCTGCTC ATTACCTCGC TTCCCGTTTT
 51  CGCCAGTGTT TTTACCGCCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA
101  AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CCTTGTCGAT
151  TTGGACAACC GCCTGACCGG ACGGCTGAAA AACATCATCA CCACCGTCGC
```

```
 201   CCTGTTCACC CTCTCCTCGC TCACGGCACA AAGCACCCTC GGCACAGGGC
 251   TGCCCTTCAT CCTCGCCATG ACCCTGATGA CCTTCGGCTT CACCATTTTA
 301   GGCGCGGTCG GGCTCAAATA CCGCACCTTC GCCTTCGGTG CACTCGCCGT
 351   CGCCACCTAC ACCACACTTA CCTACACCCC CGAAACCTAC TGGCTGACCA
 401   ACCCCTTCAT GATTTTATGC GGCACCGTAC TGTACAGCAC CGCCATCCTC
 451   CTGTTCCAAA TCGTCCTGCC CCACCGCCCC GTCCAAGAAA GCGTCGCCAA
 501   CGCCTACGAC GCACTCGGCG GCTACCTCGA AGCCAAAGCC GACTTCTTCG
 551   ACCCCGATGA GGCAGCCTGG ATAGGCAACC GCCACATCGA CCTCGCCATG
 601   AGCAACACCG GCGTCATCAC CGCCTTCAAC CAATGCCGTT CCGCCCTGTT
 651   TTACCGCCTT CGCGGCAAAC ACCGCCACCC GCGCACCGCC AAAATGCTGC
 701   GTTACTACTT TGCCGCCCAA GACATACACG AACGCATCAG CTCCGCCCAC
 751   GTCGATTATC AGGAAATGTC CGAAAAATTC AAAAACACCG ACATCATCTT
 801   CCGCATCCAC CGCCTGCTCG AAATGCAGGG ACAAGCCTGC CGCAACACCG
 851   CCCAAGCCCT GCGCGCAAGC AAAGACTACG TTTACAGCAA ACGCCTCGGC
 901   CGCGCCATCG AAGGCTGCCG CCAATCGCTG CGCCTCCTTT CAGACAGCAA
 951   CGACAGTCCC GACATCCGCC ACCTGCGCCG CCTTCTCGAC AACCTCGGCA
1001   GCGTCGACCA GCAGTTCCGC CAACTCCAGC ACAACGGCCT GCAGGCAGAA
1051   AACGACCGCA TGGGCGACAC CCGCATCGCC GCCCTCGAAA CCAGCAGCCT
1101   CAAAAACACC TGGCAGGCAA TCCGTCCGCA GCTAAACCTC GAATCAGGCG
1151   TATTCCGCCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCCTGCACC
1201   ATCGTCGAAG CCCTCAACCT CAACCTCGGC TACTGGATAC TACTGACCGC
1251   CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAAGC CGCGTCCGCC
1301   AGCGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC
1351   TACTTCACCC CGTCTGTCGA AACCAAACTC TGGATTGTCA TCGCCAGTAC
1401   CACCCTCTTT TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT
1451   TCATTACCAT TCAAGCCCTG ACCAGCCTCT CCCTCGCAGG TTTGGACGTA
1501   TACGCCGCCA TGCCCGTACG CATCATCGAC ACCATTATCG GCGCATCCCT
1551   TGCCTGGGCG GCAGTCAGCT ACCTGTGGCC AGACTGGAAA TACCTCACGC
1601   TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAACGGTGC CTATCTCGAA
1651   AAAATCACCG AACGCCTCAA AAGCGGCGAA ACCGGCGACG ACGTCGAATA
1701   CCGCGCCACC CGCCGCCGCG CCCACGAACA CACCGCCGCC CTCAGCAGCA
1751   CCCTTTCCGA CATGAGCAGC GAACCCGCAA AATTCGCCGA CAGCCTGCAA
1801   CCCGGCTTTA CCCTGCTCAA AACCGGCTAC GCCCTGACCG GCTACATCTC
1851   CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT
1901   TTACCGCACA GTTCCACCTC GCCGCCGAAC ACACCGCCCA CATCTTCCAA
1951   CACCTGCCCG AAACCGAACC CGACGACTTT CAGACAGCAC TGGATACACT
2001   GCGCGGCGAA CTCGACACCC TCCGCACCCA CAGCAGCGGA ACACAAAGCC
2051   ACATCCTCCT CCAACAGCTC CAACTCATCG CCCGACAGCT CGAACCCTAC
2101   TACCGCGCCT ACCGCCAAAT TCCGCACAGG CAGCCCCAAA ATGCAGCCTG
2151   A
```

This corresponds to the amino acid sequence <SEQ ID 106; ORF19-1>:

```
  1   MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
 51   LDNRLTGRLK NIITTVALFT LSSLTAQSTL GTGLPFILAM TLMTFGFTIL
101   GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAIL
151   LFQIVLPHRP VQESVANAYD ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM
201   SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYYFAAQ DIHERISSAH
251   VDYQEMSEKF KNTDIIFRIH RLLEMQGQAC RNTAQALRAS KDYVYSKRLG
301   RAIEGCRQSL RLLSDSNDSP DIRHLRRLLD NLGSVDQQFR QLQHNGLQAE
351   NDRMGDTRIA ALETSSLKNT WQAIRPQLNL ESGVFRHAVR LSLVVAAACT
401   IVEALNLNLG YWILLTALFV CQPNYTATKS RVRQRIAGTV LGVIVGSLVP
451   YFTPSVETKL WIVIASTTLF FMTRTYKYSF STFFITIQAL TSLSLAGLDV
501   YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLERTAAL AVCSNGAYLE
551   KITERLKSGE TGDDVEYRAT RRRAHEHTAA LSSTLSDMSS EPAKFADSLQ
601   PGFTLLKTGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ
651   HLPETEPDDF QTALDTLRGE LDTLRTHSSG TQSHILLQQL QLIARQLEPY
701   YRAYRQIPHR QPQNAA*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with predicted transmenbrane protein YHFK of *H. influenzae* (accession number P44289)

ORF19 and YHFK proteins show 45% aa identity in 97 aa overlap:

```
orf19  6    LKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLKNIITT 65
            L   +I+++PVF +V  AA  +W       +MP +LGIIAGGLVDLDN  TGRLKN+  T
YHFK   5    LNAKVISTIPVFIAVNIAAVGIWFFDISSQSMPLILGIIAGGLVDLDNRLTGRLKNVFFT 64
```

```
orf19  66  VALFTLSSLTAQSTLGTGLPFILAMTLMTXXFTILGA 102
           +  F++SS   Q  +G  + +I+ MT++T  FT++GA
YHFK   65  LIAFSISSFIVQLHIGKPIQYIVLMTVLTFIFTMIGA 101
```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF19 shows 92.2% identity over a 102aa overlap with an ORF (ORF19a) from strain A of *N. meningitidis:*

```
                     10        20        30        40        50        60
orf19.pep   MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLK
            |||| |||||||||||||||||||||||||||||||||||||||||||||   |||||
orf19a      MKTPPLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK
                     10        20        30        40        50        60

                     70        80        90       100
orf19.pep   NIITTVALFTLSSLTAQSTLGTGLPFILAMTLMTXXFTILGAX
            |||:||||||||||:|||||||||||||||||||  |||:||
orf19a      NIIATVALFTLSSLVAQSTLGTGLPFILAMTLMTFGFTIMGAVGLKYRTFAFGALAVATY
                     70        80        90       100       110       120

orf19a      TTLTYTPETYWLTNPFMILCGTVLYSTAIILFQIILPHRPVQENVANAYEALGSYLEAKA
                    130       140       150       160       170       180
```

The complete length ORF19a nucleotide sequence <SEQ ID 107> is:

```
   1  ATGAAAACCC CACCCCTCAA GCCTCTGCTC ATTACCTCGC TTCCCGTTTT
  51  CGCCAGTGTC TTTACCGCCG CCTCCATCGT CTGGCAGCTG GGCGAACCCA
 101  AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCTGGCGG CCTGGTCGAT
 151  TTGGACAACC GCCTGACCGG ACGGCTGAAA AACATCATCG CCACCGTCGC
 201  CCTGTTCACC CTCTCCTCAC TTGTCGCGCA AAGCACCCTC GGCACAGGTT
 251  TGCCATTCAT CCTCGCCATG ACCCTGATGA CTTTCGGCTT TACCATCATG
 301  GGCGCGGTCG GGCTGAAATA CCGCACCTTC GCCTTCGGCG CACTCGCCGT
 351  CGCCACCTAC ACCACACTTA CCTACACCCC CGAAACCTAC TGGCTGACCA
 401  ACCCCTTTAT GATTCTGTGC GGAACCGTAC TGTACAGCAC CGCCATCATC
 451  CTGTTCCAAA TCATCCTGCC CCACCGCCCC GTTCAAGAAA ACGTCGCCAA
 501  CGCCTACGAA GCACTCGGCA GCTACCTCGA AGCCAAAGCC GACTTTTTCG
 551  ATCCCGACGA AGCCGAATGG ATAGGCAACC GCCACATCGA CCTCGCCATG
 601  AGCAACACCG GCGTCATCAC CGCCTTCAAC CAATGCCGTT CCGCCCTGTT
 651  TTACCGCCTT CGCGGCAAAC ACCGCCACCC GCGCACCGCC AAAATGCTGC
 701  GCTACTACTT CGCCGCCCAA GACATACACG AACGCATCAG CTCCGCCCAC
 751  GTCGACTACC AAGAGATGTC CGAAAAATTC AAAAACACCG ACATCATCTT
 801  CCGCATCCAC CGCCTGCTCG AAATGCAGGG ACAAGCCTGC CGCAACACCG
 851  CCCAAGCCCT GCGCGCAAGC AAAGACTACG TTTACAGCAA ACGCCTCGGC
 901  CGCGCCATCG AAGGCTGCCG CCAATCGCTG CGCCTCCTTT CAGACAGCAA
 951  CGACAATCCC GACATCCGCC ACCTGCGCCG CCTTCTCGAC AACCTCGGCA
1001  GCGTCGACCA GCAGTTCCGC CAACTCCAGC ACAACGGCCT GCAGGCAGAA
1051  AACGACCGCA TGGGCGACAC CCGCATCGCC GCCCTCGAAA CCGGCAGCCT
1101  CAAAAACACC TGGCAGGCAA TCCGTCCGCA GCTAAACCTC GAATCAGGCG
1151  TATTCCGCCA TGCCGTCCGC CTGTCCCTTG TCGTTGCCGC CGCCTGCACC
1201  ATCGTCGAAG CCCTCAACCT CAACCTCGGC TACTGGATAC TACTGACCGC
1251  CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAAGC CGCGTCCGCC
1301  AGCGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC
1351  TACTTTACCC CCTCCGTCGA AACCAAACTC TGGATCGTCA TCGCCAGTAC
1401  CACCCTCTTT TTCATGACCC GCACCTACAA ATACAGCTTC TCGACATTTT
1451  TCATCACCAT TCAAGCCCTG ACCAGCCTCT CCCTCGCAGG GTTGGACGTA
1501  TACGCCGCCA TGCCCGTACG CATCATCGAC ACCATTATCG GCGCATCCCT
1551  TGCCTGGGCG GCAGTCAGCT ACCTGTGGCC AGACTGGAAA TACCTCACGC
1601  TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAACGGCGC CTATCTCGAA
1651  AAAATCACCG AACGCCTCAA AAGCGGCGAA ACCGGCGACG ACGTCGAATA
1701  CCGCGCCACC CGCCGCCGCG CCCACGAACA CACCGCCGCC CTCAGCAGCA
1751  CCCTTTCCGA CATGAGCAGC GAACCCGCAA AATTCGCCGA CAGCCTGCAA
1801  CCCGGCTTTA CCCTGCTCAA AACCGGCTAC GCCCTGACCG GCTACATCTC
1851  CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT
1901  TTACCGCACA GTTCCACCTC GCCGCCGAAC ACACCGCCCA CATCTTCCAA
1951  CACCTGCCCG AAACCGAACC CGACGACTTT CAGACAGCAC TGGATACACT
2001  GCGCGGCGAA CTCGACACCC TCCGCACCCA CAGCAGCGGA ACACAAAGCC
2051  ACATCCTCCT CCAACAGCTC CAACTCATCG CCCGGCAGCT CGAACCCTAC
2101  TACCGCGCCT ACCGACAAAT TCCGCACAGG CAGCCCCAAA ACGCAGCCTG
2151  A
```

This encodes a protein having amino acid sequence <SEQ ID 108>:

```
  1 MKTPPLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
 51 LDNRLTGRLK NIIATVALFT LSSLVAQSTL GTGLPFILAM TLMTFGFTIM
101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAII
151 LFQIILPHRP VQENVANAYE ALGSYLEAKA DFFDPDEAEW IGNRHIDLAM
201 SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYYFAAQ DIHERISSAH
251 VDYQEMSEKF KNTDIIFRIH RLLEMQGQAC RNTAQALRAS KDYVYSKRLG
301 RAIEGCRQSL RLLSDSNDNP DIRHLRRLLD NLGSVDQQFR QLQHNGLQAE
351 NDRMGDTRIA ALETGSLKNT WQAIRPQLNL ESGVFRHAVR LSLVVAAACT
401 IVEALNLNLG YWILLTALFV CQPNYTATKS RVRQRIAGTV LGVIVGSLVP
451 YFTPSVETKL WIVIASTTLF FMTRTYKYSF STFFITIQAL TSLSLAGLDV
501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLERTAAL AVCSNGAYLE
551 KITERLKSGE TGDDVEYRAT RRRAHEHTAA LSSTLSDMSS EPAKFADSLQ
601 PGFTLLKTGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ
651 HLPETEPDDF QTALDTLRGE LDTLRTHSSG TQSHILLQQL QLIARQLEPY
701 YRAYRQIPHR QPQNAA*
```

ORF19a and ORF19-1 show 98.3% identity in 716 aa overlap:

```
                   10        20        30        40        50        60
orf19a.pep  MKTPPLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK
            |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf19-1     MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK
                   10        20        30        40        50        60

                   70        80        90       100       110       120
orf19a.pep  NIIATVALFTLSSLVAQSTLGTGLPFILAMTLMTFGFTIMGAVGLKYRTFAFGALAVATY
            |||:||||||||||:||||||||||||||||||||||||:||||||||||||||||||||
orf19-1     NIITTVALFTLSSLTAQSTLGTGLPFILAMTLMTFGFTILGAVGLKYRTFAFGALAVATY
                   70        80        90       100       110       120

                  130       140       150       160       170       180
orf19a.pep  TTLTYTPETYWLTNPFMILCGTVLYSTAIILFQIILPHRPVQENVANAYEALGSYLEAKA
            |||||||||||||||||||||||||||||:||||:||||||||:||||||:|||:|||||
orf19-1     TTLTYTPETYWLTNPFMILCGTVLYSTAILLFQIVLPHRPVQESVANAYDALGGYLEAKA
                  130       140       150       160       170       180

                  190       200       210       220       230       240
orf19a.pep  DFFDPDEAEWIGNRHIDLAMSNTGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYYFAAQ
            |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
orf19-1     DFFDPDEAAWIGNRHIDLAMSNTGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYYFAAQ
                  190       200       210       220       230       240

                  250       260       270       280       290       300
orf19a.pep  DIHERISSAHVDYQEMSEKFKNTDIIFRIHRLLEMQGQACRNTAQALRASKDYVYSKRLG
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf19-1     DIHERISSAHVDYQEMSEKFKNTDIIFRIHRLLEMQGQACRNTAQALRASKDYVYSKRLG
                  250       260       270       280       290       300

                  310       320       330       340       350       360
orf19a.pep  RAIEGCRQSLRLLSDSNDNPDIRHLRRLLDNLGSVDQQFRQLQHNGLQAENDRMGDTRIA
            ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
orf19-1     RAIEGCRQSLRLLSDSNDSPDIRHLRRLLDNLGSVDQQFRQLQHNGLQAENDRMGDTRIA
                  310       320       330       340       350       360

                  370       380       390       400       410       420
orf19a.pep  ALETGSLKNTWQAIRPQLNLESGVFRHAVRLSLVVAAACTIVEALNLNLGYWILLTALFV
            ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf19-1     ALETSSLKNTWQAIRPQLNLESGVFRHAVRLSLVVAAACTIVEALNLNLGYWILLTALFV
                  370       380       390       400       410       420

                  430       440       450       460       470       480
orf19a.pep  CQPNYTATKSRVRQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf19-1     CQPNYTATKSRVRQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF
                430       440       450       460       470       480

                  490       500       510       520       530       540
orf19a.pep  STFFITIQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLERTAAL
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
orf19-1         STFFITIQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLERTAAL
                        490       500       510       520       530       540

                        550       560       570       580       590       600
orf19a.pep      AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSDMSSEPAKFADSLQ
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf19-1         AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSDMSSEPAKFADSLQ
                        550       560       570       580       590       600

                        610       620       630       640       650       660
orf19a.pep      PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLAAEHTAHIFQHLPETEPDDF
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf19-1         PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLAAEHTAHIFQHLPETEPDDF
                        610       620       630       640       650       660

                        670       680       690       700       710
orf19a.pep      QTALDTLRGELDTLRTHSSGTQSHILLQQLQLIARQLEPYYRAYRQIPHRQPQNAAX
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf19-1         QTALDTLRGELDTLRTHSSGTQSHILLQQLQLIARQLEPYYRAYRQIPHRQPQNAAX
                        670       680       690       700       710
```

<u>Homology with a predicted ORF from *N.gonorrhoeae*</u>

ORF19 shows 95.1% identity over a 102aa overlap with a predicted ORF (ORF19.ng) from *N. gonorrhoeae*:

```
orf19.pep     MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNXXTGRLK     60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||  |||||
orf19ng       MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK     60

orf19.pep     NIITTVALFTLSSLTAQSTLGTGLPFILAMTLMTXXFTILGAX                     103
              |||:||||||||||||||||||||||||||||||  ||||||
orf19ng       NIIATVALFTLSSLTAQSTLGTGLPFILAMTLMTFGFTILGAVGLKYRTFAFGALAVATY    120
```

An ORF19ng nucleotide sequence <SEQ ID 109> is predicted to encode a protein having amino acid sequence <SEQ ID 110>:

```
  1  MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
 51  LDNRLTGRLK NIIATVALFT LSSLTAQSTL GTGLPFILAM TLMTFGFTIL
101  GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAII
151  LFQIILPHRP VQESVANAYE ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM
201  SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYYFAAQ DIHERISSAH
251  VDYQEMSEKF KNTDIIFRIR RLLEMQGQAC RNTAQAIRSG KDYVYSKRLG
301  RAIEGCRQSL RLLSDGNDSP DIRHLSRLLD NLGSVDQQFR QLRHSDSPAE
351  NDRMGDTRIA ALETGSFKNT *
```

Further work revealed the complete nucleotide sequence <SEQ ID 111>:

```
  1  ATGAAAACCC CACTCCTCAA GCCTCTGCTC ATTACCTCGC TTCCCGTTTT
 51  CGCCAGTGTC TTTACCGCCG CCTCCATCGT CTGGCAGCTA GGCGAACCCA
101  AGCTCGCCAT GCCCTTCGTA CTCGGCATCA TCGCCGGCGG CCTGGTCGAT
151  TTGGACAACC GCCTGACCGG ACGGCTGAAA AACATCATCG CCACCGTCGC
201  CCTGTTTACC CTCTCCTCGC TCACGGCGCA AAGCACCCTC GGCACAGGGC
251  TGCCCTTCAT CCTCGCCATG ACCCTGATGA CCTTCGGCTT TACCATTTTA
301  GGCGCGGTCG GGCTGAAATA CCGCACCTTC GCCTTCGGCG CACTCGCCGT
351  CGCCACCTAC ACCACGCTTA CCTACACCCC CGAAACCTAC TGGCTGACCA
401  ACCCCTTCAT GATTTTATGC GGCACCGTAC TGTACAGCAC CGCCATCATC
451  CTGTTCCAAA TCATCCTGCC CCACCGCCCC GTCCAAGAAA GCGTCGCCAA
501  TGCCTACGAA GCACTCGGCG GCTACCTCGA AGCCAAAGCC GACTTCTTCG
551  ACCCCGATGA GGCAGCCTGG ATAGGCAACC GCCACATCGA CCTCGCCATG
601  AGCAACACCG GCGTCATCAC CGCCTTCAAC CAATGCCGTT CCGCCCTGTT
651  TTACCGTTTG CGCGGCAAAC ACCGCCACCC GCGCACCGCC AAAATGCTGC
701  GCTACTACTT CGCCGCCCAA GACATCCACG AACGCATCAG CTCCGCCCAC
751  GTCGACTACC AAGAGATGTC CGAAAAATTC AAAAACACCG ACATCATCTT
801  CCGCATCCGC CGCCTGCTCG AAATGCAGGG GCAGGCGTGC CGCAACACCG
851  CCCAAGCCAT CCGGTCGGGC AAAGACTAcg tTTACAGCAA ACGCCTCGGA
901  CGCGCCATcg aaggctgCCG CCAGTCGCTg cgcctCCTTt cagacggcaA
951  CGACAGTCCC GACATCCGCC ACCTGAGccg CCTTCTCGAC AACCTCGGca
```

```
1001 GCGTcgacca gcagtTCcgc caactCCGAC ACAgcgactC CCCCGCcgaa
1051 Aacgaccgca tgggcgacaC CCGCATCGCC GCCCtcgaaa ccggcagctT
1101 caaaaaCAcc tggcaggCAA TCCGTCCGCa gctgaaCCTC GAATCatgCG
1151 TATTCCGCCA TGCCGTCCGC CTGTCCCTCG TCGTTGCCGC CGCCTGCACC
1201 ATCGTCgaag cCCTCAACCT CAACCTCGGC TACTGGATAC TGCTGACCGC
1251 CCTTTTCGTC TGCCAACCCA ACTACACCGC CACCAAAAGC CGCGTGTACC
1301 AACGCATCGC CGGCACCGTA CTCGGCGTAA TCGTCGGCTC GCTCGTCCCC
1351 TACTTCACCC CCTCCGTCGA AACCAAACTC TGGATTGTCA TCGCCGGTAC
1401 CACCCTGTTC TTCATGACCC GCACCTACAA ATACAGTTTC TCCACCTTCT
1451 TCATCACCAT TCAGGCACTG ACCAGCCTCT CCCTCGCAGG TTTGGACGTA
1501 TACGCCGCCA TGCCCGTGCG CATCATcgaC ACCATTATCG GCGCATCCCT
1551 TGCCTGGGCG GCGGTCAGCT ACCTGTGGCC AGACTGGAAA TACCTCACGC
1601 TCGAACGCAC CGCCGCCCTT GCCGTATGCA GCAGCGGCAC ATACCTCCAA
1651 AAAATTGCCG AACGCCTCAA AACCGGCGAA ACCGGCGACG ACATAGAATA
1701 CCGCATCACC CGCCGCCGCG CCCACGAACA CACCGCCGCC CTCAGCAGCA
1751 CCCTTTCCGA CATGAGCAGC GAACCCGCAA AATTCGCCGA CAGCCTGCAA
1801 CCCGGCTTTA CCCTGCTCAA AACCGGCTAC GCCCTGACCG GCTACATCTC
1851 CGCCCTCGGC GCATACCGCA GCGAAATGCA CGAAGAATGC AGCCCCGACT
1901 TTACCGCACA GTTCCACCTT GCCGCCGAAC ACACCGCCCA CATCTTCCAA
1951 CACCTGCCCG ACATGGGACC CGACGACTTT CAGACGGCAT TGGATACACT
2001 GCGCGGCGAA CTCGGCACCC TCCGCACCCG CAGCAGCGGA ACACAAAGCC
2051 ACATCCTCCT CCAACAGCTC CAACTCATCG CccgGCAACT CGAACCCTAC
2101 TACCGCGCCT ACCGACAAAT TCCGCACAGG CAGCCCCAAA ACGCAGCCTG
2151 A
```

This corresponds to the amino acid sequence <SEQ ID 112; ORF19ng-1>:

```
  1 MKTPLLKPLL ITSLPVFASV FTAASIVWQL GEPKLAMPFV LGIIAGGLVD
 51 LDNRLTGRLK NIIATVALFT LSSLTAQSTL GTGLPFILAM TLMTFGFTIL
101 GAVGLKYRTF AFGALAVATY TTLTYTPETY WLTNPFMILC GTVLYSTAII
151 LFQIILPHRP VQESVANAYE ALGGYLEAKA DFFDPDEAAW IGNRHIDLAM
201 SNTGVITAFN QCRSALFYRL RGKHRHPRTA KMLRYYFAAQ DIHERISSAH
251 VDYQEMSEKF KNTDIIFRIR RLLEMQGQAC RNTAQAIRSG KDYVYSKRLG
301 RAIEGCRQSL RLLSDGNDSP DIRHLSRLLD NLGSVDQQFR QLRHSDSPAE
351 NDRMGDTRIA ALETGSFKNT WQAIRPQLNL ESCVFRHAVR LSLVVAAACT
401 IVEALNLNLG YWILLTALFV CQPNYTATKS RVYQRIAGTV LGVIVGSLVP
451 YFTPSVETKL WIVIAGTTLF FMTRTYKYSF STFFITIQAL TSLSLAGLDV
501 YAAMPVRIID TIIGASLAWA AVSYLWPDWK YLTLERTAAL AVCSSGTYLQ
551 KIAERLKTGE TGDDIEYRIT RRRAHEHTAA LSSTLSDMSS EPAKFADSLQ
601 PGFTLLKTGY ALTGYISALG AYRSEMHEEC SPDFTAQFHL AAEHTAHIFQ
651 HLPDMGPDDF QTALDTLRGE LGTLRTRSSG TQSHILLQQL QLIARQLEPY
701 YRAYRQIPHR QPQNAA*
```

ORF19ng-1 and ORF19-1 show 95.5% identity in 716 aa overlap:

```
                     10        20        30        40        50        60
orf19-1.pep  MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf19ng-1    MKTPLLKPLLITSLPVFASVFTAASIVWQLGEPKLAMPFVLGIIAGGLVDLDNRLTGRLK
                     10        20        30        40        50        60

                     70        80        90       100       110       120
orf19-1.pep  NIITTVALFTLSSLTAQSTLGTGLPFILAMTLMTFGFTILGAVGLKYRTFAFGALAVATY
             |||:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf19ng-1    NIIATVALFTLSSLTAQSTLGTGLPFILAMTLMTFGFTILGAVGLKYRTFAFGALAVATY
                     70        80        90       100       110       120

                    130       140       150       160       170       180
orf19-1.pep  TTLTYTPETYWLTNPFMILCGTVLYSTAILLFQIVLPHRPVQESVANAYDALGGYLEAKA
             ||||||||||||||||||||||||||||||:||||:|||||||||||||||:||||||||||
orf19ng-1    TTLTYTPETYWLTNPFMILCGTVLYSTAIILFQIILPHRPVQESVANAYEALGGYLEAKA
                    130       140       150       160       170       180

                    190       200       210       220       230       240
orf19-1.pep  DFFDPDEAAWIGNRHIDLAMSNTGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYYFAAQ
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf19ng-1    DFFDPDEAAWIGNRHIDLAMSNTGVITAFNQCRSALFYRLRGKHRHPRTAKMLRYYFAAQ
                    190       200       210       220       230       240

                    250       260       270       280       290       300
orf19-1.pep  DIHERISSAHVDYQEMSEKFKNTDIIFRIHRLLEMQGQACRNTAQALRASKDYVYSKRLG
```

```
              ||||||||||||||||||||||||||||||:|||||||||||||||||:|::||||||||||
orf19ng-1     DIHERISSAHVDYQEMSEKFKNTDIIFRIRRLLEMQGQACRNTAQAIRSGKDYVYSKRLG
                        250       260       270       280       290       300

                        310       320       330       340       350       360
orf19-1.pep   RAIEGCRQSLRLLSDSNDSPDIRHLRRLLDNLGSVDQQFRQLQHNGLQAENDRMGDTRIA
              |||||||||||||||:||||||||| ||||||||||||||||:|:   ||||||||||||
orf19ng-1     RAIEGCRQSLRLLSDGNDSPDIRHLSRLLDNLGSVDQQFRQLRHSDSPAENDRMGDTRIA
                        310       320       330       340       350       360

                        370       380       390       400       410       420
orf19-1.pep   ALETSSLKNTWQAIRPQLNLESGVFRHAVRLSLVVAAACTIVEALNLNLGYWILLTALFV
              ||||:|:||||||||||||||| |||||||||||||||||||||||||||||||||||||
orf19ng-1     ALETGSFKNTWQAIRPQLNLESCVFRHAVRLSLVVAAACTIVEALNLNLGYWILLTALFV
                        370       380       390       400       410       420

                        430       440       450       460       470       480
orf19-1.pep   CQPNYTATKSRVRQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIASTTLFFMTRTYKYSF
              |||||||||||| ||||||||||||||||||||||||||||||||:||||||||||||||
orf19ng-1     CQPNYTATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIAGTTLFFMTRTYKYSF
                        430       440       450       460       470       480

                        490       500       510       520       530       540
orf19-1.pep   STFFITIQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLERTAAL
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf19ng-1     STFFITIQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLERTAAL
                        490       500       510       520       530       540

                        550       560       570       580       590       600
orf19-1.pep   AVCSNGAYLEKITERLKSGETGDDVEYRATRRRAHEHTAALSSTLSDMSSEPAKFADSLQ
              ||||:|:||:||:||||:||||||:||| |||||||||||||||||||||||||||||||
orf19ng-1     AVCSSGTYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFADSLQ
                        550       560       570       580       590       600

                        610       620       630       640       650       660
orf19-1.pep   PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLAAEHTAHIFQHLPETEPDDF
              |||||||||||||||||||||||||||||||||||||||||||||||||||||:  ||||
orf19ng-1     PGFTLLKTGYALTGYISALGAYRSEMHEECSPDFTAQFHLAAEHTAHIFQHLPDMGPDDF
                        610       620       630       640       650       660

                        670       680       690       700       710
orf19-1.pep   QTALDTLRGELDTLRTHSSGTQSHILLQQLQLIARQLEPYYRAYRQIPHRQPQNAAX
              ||||||||||| ||||:||||||||||||||||||||||||||||||||||||||||
orf19ng-1     QTALDTLRGELGTLRTRSSGTQSHILLQQLQLIARQLEPYYRAYRQIPHRQPQNAAX
                        670       680       690       700       710
```

In addition, ORF19ng-1 shows significant homology to a hypothetical gonococcal protein previously entered in the databases:

```
sp|O33369|YOR2_NEIGO  HYPOTHETICAL  45.5  KD  PROTEIN  (ORF2)  gnl|PID|e1154438
(AJ002423) hypothetical protein [Neisseria gonorrh] Length = 417
 Score = 1512 (705.6 bits), Expect = 5.3e-203, P = 5.3e-203
 Identities = 301/326 (92%), Positives = 306/326 (93%)

Query:   307 RQSLRLLSDGNDSPDIRHLSRLLDNLGSVDQQFRQLRHSDSPAENDRMGDTRIAALETGS 366
             RQSLRLLSDGNDS DIRHLSRLLDNLGSVDQQFRQLRHSDSPAENDRMGDTRIAALETGS
Sbjct:     1 RQSLRLLSDGNDSXDIRHLSRLLDNLGSVDQQFRQLRHSDSPAENDRMGDTRIAALETGS 60

Query:   367 FKNTWQAIRPQLNLESCVFRHAVRLSLVVAAACTIVEALNLNLGYWILLTALFVCQPNYT 426
             FKNTWQAIRPQLNLES VFRHAVRLSLVVAAACTIVEALNLNLGYWILLT LFVCQPNYT
Sbjct:    61 FKNTWQAIRPQLNLESGVFRHAVRLSLVVAAACTIVEALNLNLGYWILLTRLFVCQPNYT 120

Query:   427 ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIAGTTLFFMTRTYKYSFSTFFIT 486
             ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIAGTTLFFMTRTYKYSFSTFFIT
Sbjct:   121 ATKSRVYQRIAGTVLGVIVGSLVPYFTPSVETKLWIVIAGTTLFFMTRTYKYSFSTFFIT 180

Query:   487 IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLERTAALAVCSSG 546
             IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLERTAALAVCSSG
Sbjct:   181 IQALTSLSLAGLDVYAAMPVRIIDTIIGASLAWAAVSYLWPDWKYLTLERTAALAVCSSG 240
```

```
Query:   547 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFADSLQPGFTLL 606
             TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFAD+  P
Sbjct:   241 TYLQKIAERLKTGETGDDIEYRITRRRAHEHTAALSSTLSDMSSEPAKFADTCNPALPCS 300

Query:   607 KTGYALTGYISALGAYRSEMHEECSP 632
             K   ALTGYISALG   ++   +  +P
Sbjct:   301 KPATALTGYISALGHTAAKCTKNAAP 326
```

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein (the first of which is also seen in the meningococcal protein), and on homology with the YHFK protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 14

The following DNA sequence, believed to be complete, was identified in *N.meningitidis* <SEQ ID 113>:

```
   1  ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTC
  51  GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTCGGCG
 101  CGGGTATGGC GACGGATGCG TTTTTTGTCG CGTTCAAACT GCCCAACCTG
 151  CTTCGCCGCG TGTTTGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
 201  TTTGGCGGAA TACAAGGAAA CGCGTTCAAA AGAGGCGG.C GAAGCCTTTA
 251  TCCGCCATGT GGCGGGGATG CTGTCGTTTG TACTGGTTAT CGTTACCGCG
 301  CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGAGTT
 351  TTGCCCAAGA TGCCGACAAA TTTCAGCTCT CCATCGATTT GCTGCGGATT
 401  ACGTTTCCTT ATATATTATT GATTTCCCTG TCTTCATTTG TCGGCTCGGT
 451  ACTCAATTCT TATCATAAGT TCGGCATTCC GGCGTTTACG CCAC.GTTTC
 501  TGAACGTGTC GTTTATCGTA TTCGCGCTGT TTTTCGTGCC GTATTTCGAT
 551  CCGCCCGTTA CCGCGCyGGC GTGGGCGGTC TTTGTCGGCG GCATTTTGCA
 601  ACTCGrmTTC CAACTGCCCT GGCTGGCGAA ACTGGGCTTT TTGAAACTGC
 651  CCAAACtGAG TTTCAAAGAT GCGGCGGTCA ACCGCGTGAT GAAACAGATG
 701  GCGCCTGCgA TTTTgGGCGT GAgCGTGGCG CAGGTTTCTT TGGTGATCAA
 751  CACGATTTTc GCGTCTTATC TGCAATCGGG CAGCGTTTCA TGGATGTATT
 801  ACGCCGACCG CATGATGGAG CTGCCCAGCG GCGTGCTGGG GGCGGCACTC
 851  GGTACGATTT TGCTGCCGAC TTTGTCCAAA CACTCGGCAA ACCaAGATAC
 901  GGaACAGTTT TCCGCCCTGC TCGACTGGGG TTTGCGCCTG TGCATGCtgc
 951  TGACGCTGCC GGCGgcGGTC GGACTGGCGG TGTTGTCGTT cCCgCtGGTG
1001  GCGACGCTGT TTATGTACCG CGwATTTACG CTGTTTGACG CGCAGATGAC
1051  GCAACACGCG CTGATTGCCT ATTCTTTCGG TTTAATCGGC TTAATCATGA
1101  TTAAAGTGTT GGCACCCGGC TTCTATGCGC GGCAAAACAT CAAwAmGCCC
1151  GTCAAAATCG CCATCTTCAC GCTCATCTGC mCGCAGTTGA TGAACCTTGs
1201  CTTTAyCGGC CCACTrrAAC rCasTCGGAC TTTCGCTTGC CATCGGTCTG
1251  GGCGCGTGTA TCAATGCCGG ATTGTTGTTT TACCTGTTGC GCAGACACGG
1301  TATTTACCAA CCTGG.CAAG GGTTGGGCAG CGTTCTT.AG CAAAAATGCT
1351  GcTCTCGCTC GCCGTGA
```

This corresponds to the amino acid sequence <SEQ ID 114; ORF20>:

```
   1  MNMLGALAKV GSLTMVSRVL GFVRDTVIAR AFGAGMATDA FFVAFKLPNL
  51  LRRVFAEGAF AQAFVPILAE YKETRSKEAX EAFIRHVAGM LSFVLVIVTA
 101  LGILAAPWVI YVSAPSFAQD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
 151  LNSYHKFGIP AFTPXFLNVS FIVFALFFVP YFDPPVTAXA WAVFVGGILQ
 201  LXFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQVSLVIN
 251  TIFASYLQSG SVSWMYYADR MMELPSGVLG AALGTILLPT LSKHSANQDT
 301  EQFSALLDWG LRLCMLLTLP AAVGLAVLSF PLVATLFMYR XFTLFDAQMT
 351  QHALIAYSFG LIGLIMIKVL APGFYARQNI XXPVKIAIFT LICXQLMNLX
 401  FXGPLXXIGL SLAIGLGACI NAGLLFYLLR RHGIYQPXQG LGSVLXQKCC
 451  SRSP*
```

These sequences were elaborated, and the complete DNA sequence <SEQ ID 115> is:

```
   1  ATGAATATGC TGGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTC
  51  GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTCGGCG
```

```
 101  CGGGTATGGC GACGGATGCG TTTTTTGTCG CGTTCAAACT GCCCAACCTG
 151  CTTCGCCGCG TGTTTGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
 201  TTTGGCGGAA TACAAGGAAA CGCGTTCAAA AGAGGCGGCG GAGGCTTTTA
 251  TCCGCCATGT GGCGGGGATG CTGTCGTTTG TACTGGTTAT CGTTACCGCG
 301  CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGGTTT
 351  TGCCCAAGAT GCCGACAAAT TTCAGCTCTC CATCGATTTG CTGCGGATTA
 401  CGTTTCCTTA TATATTATTG ATTTCCCTGT CTTCATTTGT CGGCTCGGTA
 451  CTCAATTCTT ATCATAAGTT CGGCATTCCG GCGTTTACGC CCACGTTTCT
 501  GAACGTGTCG TTTATCGTAT TCGCGCTGTT TTTCGTGCCG TATTTCGATC
 551  CGCCCGTTAC CGCGCTGGCG TGGGCGGTCT TTGTCGGCGG CATTTTGCAA
 601  CTCGGCTTCC AACTGCCCTG GCTGGCGAAA CTGGGCTTTT TGAAACTGCC
 651  CAAACTGAGT TTCAAAGATG CGGCGGTCAA CCGCGTGATG AAACAGATGG
 701  CGCCTGCGAT TTTGGGCGTG AGCGTGGCGC AGGTTTCTTT GGTGATCAAC
 751  ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTATTA
 801  CGCCGACCGC ATGATGGAGC TGCCCAGCGG CGTGCTGGGG GCGGCACTCG
 851  GTACGATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAAA CCAAGATACG
 901  GAACAGTTTT CCGCCCTGCT CGACTGGGGT TTGCGCCTGT GCATGCTGCT
 951  GACGCTGCCG GCGGCGGTCG GACTGGCGGT GTTGTCGTTC CCGCTGGTGG
1001  CGACGCTGTT TATGTACCGC GAATTTACGC TGTTTGACGC GCAGATGACG
1051  CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGCT TAATCATGAT
1101  TAAAGTGTTG GCACCCGGCT TCTATGCGCG GCAAAACATC AAAACGCCCG
1151  TCAAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTTGCC
1201  TTTATCGGCC CACTGAAACA CGTCGGACTT TCGCTTGCCA TCGGTCTGGG
1251  CGCGTGTATC AATGCCGGAT TGTTGTTTTA CCTGTTGCGC AGACACGGTA
1301  TTTACCAACC TGGCAAGGGT TGGGCAGCGT TCTTAGCAAA AATGCTGCTC
1351  TCGCTCGCCG TGATGTGCGG CGGACTGTGG GCAGCGCAGG CTTACCTGCC
1401  GTTTGAATGG GCGCACGCCG GCGGAATGCG GAAAGCGGGG CAGCTCTGCA
1451  TCCTGATTGC CGTCGGCGGC GGACTGTATT TCGCATCACT GGCGGCTTTG
1501  GGCTTCCGTC CGCGCCATTT CAAACGCGTG GAAAACTGA
```

This corresponds to the amino acid sequence <SEQ ID 116; ORF20-1>:

```
   1  MNMLGALAKV GSLTMVSRVL GFVRDTVIAR AFGAGMATDA FFVAFKLPNL
  51  LRRVFAEGAF AQAFVPILAE YKETRSKEAA EAFIRHVAGM LSFVLVIVTA
 101  LGILAAPWVI YVSAPGFAQD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
 151  LNSYHKFGIP AFTPTFLNVS FIVFALFFVP YFDPPVTALA WAVFVGGILQ
 201  LGFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQVSLVIN
 251  TIFASYLQSG SVSWMYYADR MMELPSGVLG AALGTILLPT LSKHSANQDT
 301  EQFSALLDWG LRLCMLLTLP AAVGLAVLSF PLVATLFMYR EFTLFDAQMT
 351  QHALIAYSFG LIGLIMIKVL APGFYARQNI KTPVKIAIFT LICTQLMNLA
 401  FIGPLKHVGL SLAIGLGACI NAGLLFYLLR RHGIYQPGKG WAAFLAKMLL
 451  SLAVMCGGLW AAQAYLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAAL
 501  GFRPRHFKRV EN*
```

Computer analysis of this amino acid sequence gave the following results:

<u>Homology with the MviN virulence factor of *S. typhimurium* (accession number P37169)</u>

ORF20 and MviN proteins show 63% aa identity in 440aa overlap:

```
Orf20  1    MNMLGALAKVGSLTMVSRVLGFVRDTVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF 60
            MN+L +LA V S+TM SRVLGF RD ++AR FGAGMATDAFFVAFKLPNLLRR+FAEGAF
MviN   14   MNLLKSLAAVSSMTMFSRVLGFARDAIVARIFGAGMATDAFFVAFKLPNLLRRIFAEGAF 73

Orf20  61   AQAFVPILAEYKETRSKEAXEAFIRHVAGMLSFVLVIVTALGILAAPWVIYVSAPSFAQD 120
            +QAFVPILAEYK  + +EA   F+ +V+G+L+  L +VT  G+LAAPWVI V+AP FA
MviN   74   SQAFVPILAEYKSKQGEEATRIFVAYVSGLLTLALAVVTVAGMLAAPWVIMVTAPGFADT 133

Orf20  121  ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFTPXFLNVSFIVFALFFVP 180
            ADKF L+  LLRITFPYILLISL+S VG++LN++++F IPAF P FLN+S I FALF  P
MviN   134  ADKFALTTQLLRITFPYILLISLASLVGAILNTWNRFSIPAFAPTFLNISMIGFALFAAP 193

Orf20  181  YFDPPVTAXAWAVFVGGILQLXFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV 240
            YF+PPV A AWAV VGG+LQL +QLP+L K+G L LP+++F+D     RV+KQM PAILGV
MviN   194  YFNPPVLALAWAVTVGGVLQLVYQLPYLKKIGMLVLPRINFRDTGAMRVVKQMGPAILGV 253

Orf20  241  SVAQVSLVINTIFASYLQSGSVSWMYYADRMMELPSGVLGAALGTILLPTLSKHSANQDT 300
            SV+Q+SL+INTIFAS+L SGSVSWMYYADR+ME PSGVLG ALGTILLP+LSK  A+ +
MviN   254  SVSQISLIINTIFASFLASGSVSWMYYADRLMEFPSGVLGVALGTILLPSLSKSFASGNH 313
```

```
Orf20  301 EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHALIAYSFG 360
           +++  L+DWGLRLC LL LP+AV L +L+ PL  +LF Y  FT FDA MTQ ALIAYS G
MviN   314 DEYCRLMDWGLRLCFLLALPSAVALGILAKPLTVSLFQYGKFTAFDAAMTQRALIAYSVG 373

Orf20  361 LIGLIMIKVLAPGFYARQNIXXPVKIAIFTLICXQLMNLXFXXXXXXXXXXXXXXXXXXCI 420
           LIGLI++KVLAPGFY+RQ+I  PVKIAI TLI  QLMNL F                  C+
MviN   374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433

Orf20  421 NAGLLFYLLRRHGIYQPXQG 440
           NA LL++ LR+  I+ P  G
MviN   434 NASLLYWQLRKQNIFTPQPG 453
```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF20 shows 93.5% identity over a 447aa overlap with an ORF (ORF20a) from strain A of *N. meningitidis:*

```
                      10        20        30        40        50        60
orf20.pep    MNMLGALAKVGSLTMVSRVLGFVRDTVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
             ||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
orf20a       MNMLGALVKVGSLTMVSRVLGFVRDTVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
                      10        20        30        40        50        60

                      70        80        90       100       110       120
orf20.pep    AQAFVPILAEYKETRSKEAXEAFIRHVAGMLSFVLVIVTALGILAAPWVIYVSAPSFAQD
             |||||||||||||||||||:|||||||||||||||||||||||||||||||||||:||:|
orf20a       AQAFVPILAEYKETRSKEATEAFIRHVAGMLSFVLVIVTALGILAAPWVIYVSAPGFAKD
                      70        80        90       100       110       120

                     130       140       150       160       170       180
orf20.pep    ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFTPXFLNVSFIVFALFFVP
             |||||||||||||||||||||||||||||||||||||:||||||:|||||||||||||||
orf20a       ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFSIPAFTPTFLNVSFIVFALFFVP
                     130       140       150       160       170       180

                     190       200       210       220       230       240
orf20.pep    YFDPPVTAXAWAVFVGGILQLXFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV
             |||||||| ||||||||||||| ||||||||||||||||||||||||||||||||||||||
orf20a       YFDPPVTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV
                     190       200       210       220       230       240

                     250       260       270       280       290       300
orf20.pep    SVAQVSLVINTIFASYLQSGSVSWMYYADRMMELPSGVLGAALGTILLPTLSKHSANQDT
             ||||:||||||||||||||||||||||||||||||:||||||||||||||||||||||||
orf20a       SVAQISLVINTIFASYLQSGSVSWMYYADRMMELPGGVLGAALGTILLPTLSKHSANQDT
                     250       260       270       280       290       300

                     310       320       330       340       350       360
orf20.pep    EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHALIAYSFG
             |||||||||||| |||||||||||:|||||||||||||||| |||||||||||||||||||
orf20a       EQFSALLDWGLRXCMLLTLPAAVGMAVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG
                     310       320       330       340       350       360

                     370       380       390       400       410       420
orf20.pep    LIGLIMIKVLAPGFYARQNIXXPVKIAIFTLICXQLMNLXFXGPLXXIGLSLAIGLGACI
             |||||||||||||||||||| :|||||||||||:||||| | |||  :||||||||||||
orf20a       LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI
                     370       380       390       400       410       420

                     430       440       450
orf20.pep    NAGLLFYLLRRHGIYQPXQGLGSVLXQKCCSRSPX
             ||||||||||||||||| :| :: | :
orf20a       NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMGGGLYAAQIWLPFDWAHAGGMQKAA
                     430       440       450       460       470       480
```

The complete length ORF20a nucleotide sequence <SEQ ID 117> is:

```
  1 ATGAATATGC TGGGAGCTTT GGTAAAAGTC GGCAGCCTGA CGATGGTGTC
 51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGC GCATTCGGCG
101 CAGGCATGGC GACGGATGCG TTCTTTGTCG CGTTCAAACT GCCCAACCTG
```

```
 151  CTTCGCCGCG TGTTTGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
 201  TTTGGCGGAA TATAAGGAAA CGCGTTCTAA AGAGGCGACG GAGGCTTTTA
 251  TCCGCCATGT GGCGGGGATG CTGTCGTTTG TACTGGTCAT CGTTACCGCG
 301  CTGGGCATAC TTGCCGCGCC TTGGGTGATT TATGTTTCCG CACCCGGTTT
 351  TGCCAAAGAT GCCGACAAAT TTCAGCTCTC TATCGATTTG CTGCGGATTA
 401  CGTTTCCTTA TATCTTATTG ATTTCACTTT CCTCTTTTGT CGGCTCGGTA
 451  CTCAATTCCT ATCATAAATT CAGCATTCCT GCGTTTACGC CCACGTTCCT
 501  GAACGTGTCG TTTATCGTAT TCGCGCTGTT TTTCGTGCCG TATTTCGATC
 551  CTCCCGTTAC CGCGCTGGCT TGGGCGGTTT TTGTCGGCGG CATTTTGCAA
 601  CTCGGCTTCC AACTGCCCTG GCTGGCGAAA CTGGGTTTTT TGAAACTGCC
 651  CAAACTGAGT TTCAAAGATG CGGCGGTCAA CCGCGTGATG AAACAGATGG
 701  CGCCTGCGAT TTTGGGCGTG AGCGTGGCGC AGATTTCTTT GGTGATCAAC
 751  ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTATTA
 801  CGCCGACCGC ATGATGGAAC TGCCCGGCGG CGTGCTGGGG GCGGCACTCG
 851  GTACGATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAAA CCAAGATACG
 901  GAACAGTTTT CCGCCCTGCT CGACTGGGGT TTGCGCNTGT GCATGCTGCT
 951  GACGCTGCCG GCGGCGGTCG GAATGGCGGT GTTGTCGTTC CCGCTGGTGG
1001  CAACCTTGTT TATGTACCGA GAATTCACGC TGTTTGACGC GCAGATGACG
1051  CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGTT TAATCATGAT
1101  TAAAGTGTTG GCGCCCGGCT TTTATGCGCG GCAAAACATC AAAACGCCCG
1151  TCAAAATCGC CATCTTCACG CTCATTTGCA CGCAGTTGAT GAACCTTGCC
1201  TTTATCGGCC CACTGAAACA CGTCGGACTT TCGCTTGCCA TCGGTCTGGG
1251  CGCGTGTATC AATGCCGGAT TGTTGTTTTA CCTGTTGCGC AGACACGGTA
1301  TTTACCAACC TGGCAAGGGT TGGGCAGCGT TCTTGGCAAA AATGCTGCTC
1351  TCGCTCGCCG TGATGGGAGG CGGCCTGTAT GCCGCCCAAA TCTGGCTGCC
1401  GTTCGACTGG GCACACGCCG GCGGAATGCA AAAGGCCGCC CGGCTCTTCA
1451  TCCTGATTGC CGTCGGCGGC GGACTGTATT TCGCATCACT GGCGGCTTTG
1501  GGCTTCCGTC CGCGCCATTT CAAACGCGTG GAAAGCTGA
```

This encodes a protein having amino acid sequence <SEQ ID 118>:

```
  1  MNMLGALVKV GSLTMVSRVL GFVRDTVIAR AFGAGMATDA FFVAFKLPNL
 51  LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLVIVTA
101  LGILAAPWVI YVSAPGFAKD ADKFQLSIDL LRITFPYILL ISLSSFVGSV
151  LNSYHKFSIP AFTPTFLNVS FIVFALFFVP YFDPPVTALA WAVFVGGILQ
201  LGFQLPWLAK LGFLKLPKLS FKDAAVNRVM KQMAPAILGV SVAQISLVIN
251  TIFASYLQSG SVSWMYYADR MMELPGGVLG AALGTILLPT LSKHSANQDT
301  EQFSALLDWG LRXCMLLTLP AAVGMAVLSF PLVATLFMYR EFTLFDAQMT
351  QHALIAYSFG LIGLIMIKVL APGFYARQNI KTPVKIAIFT LICTQLMNLA
401  FIGPLKHVGL SLAIGLGACI NAGLLFYLLR RHGIYQPGKG WAAFLAKMLL
451  SLAVMGGGLY AAQIWLPFDW AHAGGMQKAA RLFILIAVGG GLYFASLAAL
501  GFRPRHFKRV ES*
```

ORF20a and ORF20-1 show 96.5% identity in 512 aa overlap:

```
                       10        20        30        40        50        60
orf20a.pep     MNMLGALVKVGSLTMVSRVLGFVRDTVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
               |||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
orf20-1        MNMLGALAKVGSLTMVSRVLGFVRDTVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
                       10        20        30        40        50        60

                       70        80        90       100       110       120
orf20a.pep     AQAFVPILAEYKETRSKEATEAFIRHVAGMLSFVLVIVTALGILAAPWVIYVSAPGFAKD
               |||||||||||||||||||:||||||||||||||||||||||||||||||||||||||:|
orf20-1        AQAFVPILAEYKETRSKEAAEAFIRHVAGMLSFVLVIVTALGILAAPWVIYVSAPGFAQD
                       70        80        90       100       110       120

                      130       140       150       160       170       180
orf20a.pep     ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFSIPAFTPTFLNVSFIVFALFFVP
               |||||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
orf20-1        ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFG1PAFTPTFLNVSFIVFALFFVP
                      130       140       150       160       170       180

                      190       200       210       220       230       240
orf20a.pep     YFDPPVTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf20-1        YFDPPVTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV
                      190       200       210       220       230       240

                      250       260       270       280       290       300
orf20a.pep     SVAQISLVINTIFASYLQSGSVSWMYYADRMMELPGGVLGAALGTILLPTLSKHSANQDT
```

```
                ||||:||||||||||||||||||||||||||||||:||||||||||||||||||||||||
orf20-1         SVAQVSLVINTIFASYLQSGSVSWMYYADRMMELPSGVLGAALGTILLPTLSKHSANQDT
                          250       260       270       280       290       300

                          310       320       330       340       350       360
orf20a.pep      EQFSALLDWGLRXCMLLTLPAAVGMAVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG
                |||||||||||| |||||||||||:|||||||||||||||||||||||||||||||||||
orf20-1         EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG
                          310       320       330       340       350       360

                          370       380       390       400       410       420
orf20a.pep      LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf20-1         LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI
                          370       380       390       400       410       420

                          430       440       450       460       470       480
orf20a.pep      NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMGGGLYAAQIWLPFDWAHAGGMQKAA
                ||||||||||||||||||||||||||||||||||| |||:||| :|||:|||||||:||:
orf20-1         NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMCGGLWAAQAYLPFEWAHAGGMRKAG
                          430       440       450       460       470       480

                          490       500       510
orf20a.pep      RLFILIAVGGGLYFASLAALGFRPRHFKRVESX
                :| ||||||||||||||||||||||||||||:|
orf20-1         QLCILIAVGGGLYFASLAALGFRPRHFKRVENX
                          490       500       510
```

<u>Homology with a predicted ORF from *N.gonorrhoeae*</u>

ORF20 shows 92.1% identity over a 454aa overlap with a predicted ORF (ORF20ng) from *N. gonorrhoeae:*

```
orf20.pep   MNMLGALAKVGSLTMVSRVLGFVRDTVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF    60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf20ng     MNMLGALAKVGSLTMVSRVLGFVRDTVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF    60

orf20.pep   AQAFVPILAEYKETRSKEAXEAFIRHVAGMLSFVLVIVTALGILAAPWVIYVSAPSFAQD   120
            |||||||||||||||||||:|||||||||||||||::||||||||||||||||||:|::|
orf20ng     AQAFVPILAEYKETRSKEATEAFIRHVAGMLSFVLIVVTALGILAAPWVIYVSAPGFTKD   120

orf20.pep   ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFTPXFLNVSFIVFALFFVP   180
            ||||||||:||||||||||||||||||||:||||||||||||||:|||:|||||||||||
orf20ng     ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFTPTFLNISFIVFALFFVP   180

orf20.pep   YFDPPVTAXAWAVFVGGILQLXFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV   240
            |||||||| |||||||||||| |||||||||||||||||:||||||||||||||||||||
orf20ng     YFDPPVTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLNFKDAAVNRVMKQMAPAILGV   240

orf20.pep   SVAQVSLVINTIFASYLQSGSVSWMYYADRMMELPSGVLGAALGTILLPTLSKHSANQDT   300
            ||||:||||||||||||||||||||||||||||||:||||||||||||||||||||||||
orf20ng     SVAQISLVINTIFASYLQSGSVSWMYYADRMMELPGGVLGAALGTILLPTLSKHSANQDT   300

orf20.pep   EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYRXFTLFDAQMTQHALIAYSFG   360
            ||||||||||||||||||||||:||||||||||||||||| |||||||||||||||||||
orf20ng     EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG   360

orf20.pep   LIGLIMIKVLAPGFYARQNIXXPVKIAIFTLICXQLMNLXFXGPLXXIGLSLAIGLGACI   420
            ||||||||||| |||||||| :|||||||||||:||||| | |||   ||||||||||||
orf20ng     LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI   420

orf20.pep   NAGLLFYLLRRHGIYQPXQGLGSVLXQKCCSRSP   454
            ||||||:|:|:||||:| ||||:   :|||||||
orf20ng     NAGLLFFLFRKHGIYRPGQGLGQPSWRKCCSRSP   454
```

An ORF20ng nucleotide sequence <SEQ ID 119> was predicted to encode a protein having amino acid sequence <SEQ ID 120>:

```
  1 MNMLGALAKV GSLTMVSRVL GFVRDTVIAR AFGAGMATDA FFVAFKLPNL
 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLIVVTA
101 LGILAAPWVI YVSAPGFTKD ADKFQLSISL LRITFPYILL ISLSSFVGSI
151 LNSYHKFGIP AFTPTFLNIS FIVFALFFVP YFDPPVTALA WAVFVGGILQ
201 LGFQLPWLAK LGFLKLPKLN FKDAAVNRVM KQMAPAILGV SVAQISLVIN
251 TIFASYLQSG SVSWMYYADR MMELPGGVLG AALGTILLPT LSKHSANQDT
301 EQFSALLDWG LRLCMLLTLP AAAGLAVLSF PLVATLFMYR EFTLFDAQMT
351 QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNLA
401 FIGPLKHAGL SLAIGLGACI NAGLLFFLFR KHGIYRPGQG LGQPSWRKCC
451 SRSP*
```

Further DNA sequence analysis revealed the following DNA sequence <SEQ ID 121>:

```
   1 ATGAATATGC TTGGAGCTTT GGCAAAAGTC GGCAGCCTGA CGATGGTGTC
  51 GCGCGTTTTG GGATTTGTGC GCGATACGGT CATTGCGCGG GCATTCGGCG
 101 CGGGTATGGC GACGGATGCG TTTTTTGTCG CGTTCAAACT GCCCAACCTG
 151 CTTCGCCGCG TGTTTGCGGA GGGGGCGTTT GCCCAAGCGT TTGTGCCGAT
 201 TTTGGCGGAA TATAAGGAAA CGCGTTCTAA AGAGGCGAcg gAGGCTTTTA
 251 TCCGCCACGt tgcgggAatg CTGTCGTTTG TGCTGATcgt cGttacCGCG
 301 CTGGGCATAC TTGCCGCgcc tTGGGTGATT TATGTTtccg CgcccGGCTT
 351 TACCAAAGAC GCGGACAAGT TCCAACTTTC CATCAGCCTG CTGCGGATTA
 401 CGTTTCCTTA TATATTATTG ATTTCTTTGT CTTCTTTTGT CGGCTCGATA
 451 CTCAATTCCT ACCATAAGTT CGGCATTCCC GCGTTTACGC CCACGTTTTT
 501 AAACATCTCT TTTATCGTAT TCGCACTGTT TTTCGTGCCG TATTTCGATC
 551 CGCCCGTTAC CGCGCTGGCG TGGGCGGTTT TTGTCGGCGG TATTTTGCAG
 601 CTCGGTTTCC AACTGCCGTG GCTGGCGAAA CTGGGCTTTT TGAAACTGCC
 651 CAAACTGAAT TTCAAAGATG CGGCGGTCAA CCGCGTCATG AAACAGATGG
 701 CGCCTGCGAT TTTGGGCGTG agcgTGGCGC AAATTTCTTT GgttATCAAC
 751 ACGATTTTCG CGTCTTATCT GCAATCGGGC AGCGTTTCAT GGATGTatta
 801 cgCCGACCGC ATGATGGAGc tgcgccGGGG CGTGCTGGGG GCTGCACTCG
 851 GTACAATTTT GCTGCCGACT TTGTCCAAAC ACTCGGCAAA CCAAGATACG
 901 GAACAGTTTT CCGCCCTGCT CGACTGGGGT TTGCGCCTGT GCATGCTGCT
 951 GACGCTGCCG GCGGCGGccg GACTGGCGGT ATTGTCGTTC CCGCTGGTGG
1001 CGACGCTGTT TATGTACCGA GAATTCACGC TGTTTGACGC ACAAATGACG
1051 CAACACGCGC TGATTGCCTA TTCTTTCGGT TTAATCGGTT TAATTATGAT
1101 TAAAGTGTTG GCATCCGGCT TTTATGCGCG GCAAAACATC AAAACGCCCG
1151 TCAAAATCGC CATCTTCACG CTCATCTGCA CGCAGTTGAT GAACCTCGCC
1201 TTTATCGGTC CGTTGAAACA CGCCGGGCTT TCGCTCGCCA TCGGCCTGGG
1251 CGCGTGCATC AACGCCGGAT TGTTGTTCTT CCTGTTGCGC AAACACGGTA
1301 TTTACCGGCC cggcaggggt tgggcggcgt TCTTGGCGAA AATGCTGCTC
1351 GCGCTCGCCG TGATGTGCGG CGGACTGTGG GCGGCGCAGG CTTGCCTGCC
1401 GTTCGAATGG GCGCACGCCG GCGGAATGCG GAAAGCGGGG CAGCTCTGCA
1451 TCCTGATTGC CGTCGGCGGC GGACTGTATT TCGCATCTCT GGCGGCTTTG
1501 GGCTTCCGTC CGCGCCATTT CAAACGCGTG GAAAGCTGA
```

This encodes the following amino acid sequence <SEQ ID 122; ORF20ng-1>:

```
  1 MNMLGALAKV GSLTMVSRVL GFVRDTVIAR AFGAGMATDA FFVAFKLPNL
 51 LRRVFAEGAF AQAFVPILAE YKETRSKEAT EAFIRHVAGM LSFVLIVVTA
101 LGILAAPWVI YVSAPGFTKD ADKFQLSISL LRITFPYILL ISLSSFVGSI
151 LNSYHKFGIP AFTPTFLNIS FIVFALFFVP YFDPPVTALA WAVFVGGILQ
201 LGFQLPWLAK LGFLKLPKLN FKDAAVNRVM KQMAPAILGV SVAQISLVIN
251 TIFASYLQSG SVSWMYYADR MMELRRGVLG AALGTILLPT LSKHSANQDT
301 EQFSALLDWG LRLCMLLTLP AAAGLAVLSF PLVATLFMYR EFTLFDAQMT
351 QHALIAYSFG LIGLIMIKVL ASGFYARQNI KTPVKIAIFT LICTQLMNLA
401 FIGPLKHAGL SLAIGLGACI NAGLLFFLLR KHGIYRPGRG WAAFLAKMLL
451 ALAVMCGGLW AAQACLPFEW AHAGGMRKAG QLCILIAVGG GLYFASLAAL
501 GFRPRHFKRV ES*
```

ORF20ng-1 and ORF20-1 show 95.7% identity in 512 aa overlap:

```
                     10        20        30        40        50        60
orf20-1.pep  MNMLGALAKVGSLTMVSRVLGFVRDTVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf20ng-1    MNMLGALAKVGSLTMVSRVLGFVRDTVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF
                     10        20        30        40        50        60

                     70        80        90       100       110       120
orf20-1.pep  AQAFVPILAEYKETRSKEAAEAFIRHVAGMLSFVLVIVTALGILAAPWVIYVSAPGFAQD
             |||||||||||||||||||:|||||||||||||||::||||||||||||||||||||::|
orf20ng-1    AQAFVPILAEYKETRSKEATEAFIRHVAGMLSFVLIVVTALGILAAPWVIYVSAPGFTKD
```

```
                    70        80        90       100       110       120

                   130       140       150       160       170       180
orf20-1.pep ADKFQLSIDLLRITFPYILLISLSSFVGSVLNSYHKFGIPAFTPTFLNVSFIVFALFFVP
            ||||||||:|||||||||||||||||||||:||||||||||||||||||:||||||||||
orf20ng-1   ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFTPTFLNISFIVFALFFVP
                   130       140       150       160       170       180

                   190       200       210       220       230       240
orf20-1.pep YFDPPVTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLSFKDAAVNRVMKQMAPAILGV
            |||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
orf20ng-1   YFDPPVTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLNFKDAAVNRVMKQMAPAILGV
                   190       200       210       220       230       240

                   250       260       270       280       290       300
orf20-1.pep SVAQVSLVINTIFASYLQSGSVSWMYYADRMMELPSGVLGAALGTILLPTLSKHSANQDT
            ||||:||||||||||||||||||||||||||||  |||||||||||||||||||||||||
orf20ng-1   SVAQISLVINTIFASYLQSGSVSWMYYADRMMELRRGVLGAALGTILLPTLSKHSANQDT
                   250       260       270       280       290       300

                   310       320       330       340       350       360
orf20-1.pep EQFSALLDWGLRLCMLLTLPAAVGLAVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG
            ||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
orf20ng-1   EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG
                   310       320       330       340       350       360

                   370       380       390       400       410       420
orf20-1.pep LIGLIMIKVLAPGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHVGLSLAIGLGACI
            ||||||||||| ||||||||||||||||||||||||||||||||||||:|||||||||||
orf20ng-1   LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI
                   370       380       390       400       410       420

                   430       440       450       460       470       480
orf20-1.pep NAGLLFYLLRRHGIYQPGKGWAAFLAKMLLSLAVMCGGLWAAQAYLPFEWAHAGGMRKAG
            ||||||:|||:||||:||:|||||||||||:||||||||||||| |||||||||||||||
orf20ng-1   NAGLLFFLLRKHGIYRPGRGWAAFLAKMLLALAVMCGGLWAAQACLPFEWAHAGGMRKAG
                   430       440       450       460       470       480

                   490       500       510
orf20-1.pep QLCILIAVGGGLYFASLAALGFRPRHFKRVENX
            |||||||||||||||||||||||||||||||:|
orf20ng-1   QLCILIAVGGGLYFASLAALGFRPRHFKRVESX
                   490       500       510
```

In addition, ORF20ng-1 shows significant homology with a virulence factor of *S.typhimurium*:

```
sp|P37169|MVIN_SALTY VIRULENCE FACTOR MVIN pir||S40271 mviN protein - Salmonella
typhimurium gi|438252 (Z26133) mviB gene product [Salmonella typhimurium]
gnl|PID|d1005521 (D25292) ORF2 [Salmonella typhimurium] Length = 524
 Score = 1573 (750.1 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220
 Identities = 309/467 (66%), Positives = 368/467 (78%)

Query:     1 MNMLGALAKVGSLTMVSRVLGFVRDTVIARAFGAGMATDAFFVAFKLPNLLRRVFAEGAF 60
             MN+L +LA V S+TM SRVLGF RD ++AR FGAGMATDAFFVAFKLPNLLRR+FAEGAF
Sbjct:    14 MNLLKSLAAVSSMTMFSRVLGFARDAIVARIFGAGMATDAFFVAFKLPNLLRRIFAEGAF 73

Query:    61 AQAFVPILAEYKETRSKEATEAFIRHVAGMLSFVLIVVTALGILAAPWVIYVSAPGFTKD 120
             +QAFVPILAEYK  + +EAT  F+ +V+G+L+  L VVT  G+LAAPWVI V+APGF
Sbjct:    74 SQAFVPILAEYKSKQGEEATRIFVAYVSGLLTLALAVVTVAGMLAAPWVIMVTAPGFADT 133

Query:   121 ADKFQLSISLLRITFPYILLISLSSFVGSILNSYHKFGIPAFTPTFLNISFIVFALFFVP 180
             ADKF L+  LLRITFPYILLISL+S VG+ILN++++F IPAF PTFLNIS I FALF  P
Sbjct:   134 ADKFALTTQLLRITFPYILLISLASLVGAILNTWNRFSIPAFAPTFLNISMIGFALFAAP 193

Query:   181 YFDPPVTALAWAVFVGGILQLGFQLPWLAKLGFLKLPKLNFKDAAVNRVMKQMAPAILGV 240
             YF+PPV ALAWAV VGG+LQL +QLP+L K+G L LP++NF+D    RV+KQM PAILGV
Sbjct:   194 YFNPPVLALAWAVTVGGVLQLVYQLPYLKKIGMLVLPRINFRDTGAMRVVKQMGPAILGV 253

Query:   241 SVAQISLVINTIFASYLQSGSVSWMYYADRMMELRRGVLGAALGTILLPTLSKHSANQDT 300
             SV+QISL+INTIFAS+L SGSVSWMYYADR+ME   GVLG ALGTILLP+LSK  A+ +
Sbjct:   254 SVSQISLIINTIFASFLASGSVSWMYYADRLMEFPSGVLGVALGTILLPSLSKSFASGNH 313
```

```
Query:   301 EQFSALLDWGLRLCMLLTLPAAAGLAVLSFPLVATLFMYREFTLFDAQMTQHALIAYSFG 360
             +++  L+DWGLRLC LL LP+A  L +L+ PL  +LF Y +FT FDA MTQ ALIAYS G
Sbjct:   314 DEYCRLMDWGLRLCFLLALPSAVALGILAKPLTVSLFQYGKFTAFDAAMTQRALIAYSVG 373

Query:   361 LIGLIMIKVLASGFYARQNIKTPVKIAIFTLICTQLMNLAFIGPLKHAGLSLAIGLGACI 420
             LIGLI++KVLA GFY+RQ+IKTPVKIAI TLI TQLMNLAFIGPLKHAGLSL+IGL AC+
Sbjct:   374 LIGLIVVKVLAPGFYSRQDIKTPVKIAIVTLIMTQLMNLAFIGPLKHAGLSLSIGLAACL 433

Query:   421 NAGLLFFLLRKHGIYRPGRGWXXXXXXXXXXXXXVMCGGLWAAQACLP 467
             NA LL++ LRK  I+ P  GW             VM   L+      +P
Sbjct:   434 NASLLYWQLRKQNIFTPQPGWMWFLMRLIISVLVMAAVLFGVLHIMP 480

 Score = 70 (33.4 bits), Expect = 1.1e-220, Sum P(2) = 1.1e-220
 Identities = 14/41 (34%), Positives = 23/41 (56%)

Query:   469 EWAHAGGMRKAGQLCILIAVGGGLYFASLAALGFRPRHFKR 509
             EW+     + +  +L  ++  G   YFA+LA LGF+ + F R
Sbjct:   481 EWSQGSMLWRLLRLMAVVIAGIAAYFAALAVLGFKVKEFVR 521
```

Based on this analysis, including the homology with a virulence factor from *S.typhimurium*, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 15

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 123>:

```
  1 atGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGCCGTT tACGACGGCC CGGCCaTTAC CGAAGtCGCG TTGCTTGGCG
101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC
151 GTcAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT
201 GTTTACTGCG CCGGCTTCAG GcAAAATCGC CGCGATTCAC CGTGGCGAAA
251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAArGCAA CGACGAAATC
301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACTTAA GCGGCGAAGA
351 AGTGCGCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC
401 GTCCGTTCAG CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
451 GTCAATGCGA tGGACACCAA TCCG..
```

This corresponds to the amino acid sequence <SEQ ID 124; ORF22>:

```
  1 MIKIKKGLNL PIAGRPEQAV YDGPAITEVA LLGEEYAGMR PSMKVKEGDA
 51 VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLQSV VIAVEXNDEI
101 EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
151 VNAMDTNP..
```

Further work revealed the complete nucleotide sequence <SEQ ID 125>:

```
  1 ATGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
 51 GCAAGCCGTT TACGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG
101 AAGAATATGC CGGTATGCGC CCCTCGATGA AAGTCAAGGA AGGCGATGCC
151 GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTGGT
201 GTTTACTGCG CCGGCTTCAG GCAAAATCGC CGCGATTCAC CGTGGCGAAA
251 AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
301 GAGTTTGAAC GCTACGCACC TGAAGCGCTG GCAAACTTAA GCGGCGAAGA
351 AGTGCGCCGC AACCTGATCC AATCCGGTTT GTGGACTGCG CTGCGCACCC
401 GTCCGTTCAG CAAAATTCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
451 GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTCATTAT
501 CAAAGAAGCC GCCGAGGATT TCAAACGCGG CCTGTTGGTA TTGAGCCGTT
551 TGACCGAACG CAAAATCCAT GTTTGTAAGG CAGCTGGCGC AGACGTGCCG
601 TCTGAAAATG CTGCCAACAT CGAAACACAT GAATTCGGCG GCCCGCATCC
651 TGCCGGTTTG AGTGGCACGC ACATTCATTT CATCGAGCCG GTCGGCGCGA
701 ATAAAACCGT GTGGACCATC AATTATCAAG ATGTAATTAC CATTGGCCGT
751 TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CCCTAGGTGG
801 TTCTCAAGTC AACAAACCGC GCCTCTTGCG TACCGTTTTG GGTGCGAAAG
851 TATCGCAAAT TACTGCGGGC GAATTGGTTG ACACAGACAA CCGCGTGATT
901 TCCGGTTCGG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT
```

```
 951  GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
1001  AGCTGTTCGG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGT
1051  ACAACCCTCG GCCATTTCCT GAAAAACAAA CTCTTCAAGT TCAACACAGC
1101  CGTCAACGGC GGCGACCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG
1151  TGATGCCCTT GGATATCCTG CCCACCCTGC TTTTGCGCGA TTTAATCGTC
1201  GGCGATACCG ACAGCGCGCA GGCATTGGGT TGCTTGGAAT TGGACGAAGA
1251  AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATACGGCC
1301  CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA
```

This corresponds to the amino acid sequence <SEQ ID 126; ORF22-1>:

```
  1  MIKIKKGLNL PIAGRPEQAV YDGPAITEVA LLGEEYAGMR PSMKVKEGDA
 51  VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLQSV VIAVEGNDEI
101  EFERYAPEAL ANLSGEEVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
151  VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADVP
201  SENAANIETH EFGGPHPAGL SGTHIHFIEP VGANKTVWTI NYQDVITIGR
251  LFATGRLNTE RVIALGGSQV NKPRLLRTVL GAKVSQITAG ELVDTDNRVI
301  SGSVLNGAIT QGAHDYLGRY HNQISVIEEG RSKELFGWVA PQPDKYSITR
351  TTLGHFLKNK LFKFNTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV
401  GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRKVL ETIEKEG*
```

Further work identified the corresponding gene in strain A of *N.meningitidis* <SEQ ID 127>:

```
   1  ATGATTAAAA TCAAAAAAGG TCTAAACCTG CCCATCGCGG GCAGACCGGA
  51  GCAAGTCATT TATGACGGGC CCGTCATTAC CGAAGTCGCG TTGCTTGGCG
 101  AAGAATATGC CGGTATGCGC CCCTNGATGA AAGTCAAGGA AGGCGATGCC
 151  GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGNATC CGGGCGTGGT
 201  GTTTACCGCG CCNGTTTCAG GCAAAATCGC CGCCATCCAT CGCGGCGAAA
 251  AGCGCGTACT TCAGTCGGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
 301  GAGTTCGAAC GCTACGCGCC CGAAGCGTTG GCAAACTTAA GCGGCGANGA
 351  ANTNNGNNGC AATCTGATCC AATCCGGTTT GTGGACTGCG CTGCGTANCC
 401  GTCCGTTCAG CAAAATCCCT GCCGTCGATG CCGAGCCGTT CGCCATCTTC
 451  GTCAATGCGA TGGACACCAA TCCGCTNGCG GCAGACCCTG TGGTTGTGAT
 501  CAAAGAAGCC GNCGANGATT TCAGACGANG TNTGCTGGTA TTGAGCCGTT
 551  TGACCGAGCG TAAAATCCAT GTGTGTAAGG CAGCTGGCGC AGACGTGCCG
 601  TCTGAAAATG CTGCCAACAT CGAAACACAT GAATTCGGCG GCCCGCATCC
 651  GGCCGGTTTG AGTGGCACGC ACATTCATTT CATTGAGCCG GTCGGTGCAA
 701  ACAAAACCGT TTGGACCATC AATTATCAAG ATGTAATTGC CATCGGACGT
 751  TTGTTTGCAA CAGGCCGTCT GAACACCGAG CGCGTGATTG CTTTGGGTGG
 801  TTCTCAAGTC AACAAACCAC GCCTCTTGCG TACCGTTTTG GGTGCGAAAG
 851  TATCGCAAAT TACTGCGGGC GAATTGGTTG ACGCAGACAA CCGCGTGATT
 901  TCCGGTTCGG TATTGAACGG CGCGATTACA CAAGGCGCGC ACGATTATTT
 951  GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
1001  AGCTGTTCGG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGT
1051  ACGACCCTCG GCCATTTCCT GAAAAACAAA CTCTTCAAGT TCACGACAGC
1101  CGTCAACGGT GGCGACCGCG CCATGGTGCC GATTGGTACT TACGAGCGCG
1151  TAATGCCGCT AGACATCCTG CCTACCCTGC TTTTGCGCGA TTTAATCGTC
1201  GGCGATACCG ACAGCGCGCA AGCATTGGGT TGCTTGGAAT TGGACGAAGA
1251  AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATANGGCC
1301  CGCTGTTGCG TAAGGTGCTG GAAACCNTTG AGAAGGAAGG CTGA
```

This encodes a protein having amino acid sequence <SEQ ID 128; ORF22a>:

```
  1  MIKIKKGLNL PIAGRPEQVI YDGPVITEVA LLGEEYAGMR PXMKVKEGDA
 51  VKKGQVLFED KKXPGVVFTA PVSGKIAAIH RGEKRVLQSV VIAVEGNDEI
101  EFERYAPEAL ANLSGXEXXX NLIQSGLWTA LRXRPFSKIP AVDAEPFAIF
151  VNAMDTNPLA ADPVVVIKEA XXDFRRXXLV LSRLTERKIH VCKAAGADVP
201  SENAANIETH EFGGPHPAGL SGTHIHFIEP VGANKTVWTI NYQDVIAIGR
251  LFATGRLNTE RVIALGGSQV NKPRLLRTVL GAKVSQITAG ELVDADNRVI
301  SGSVLNGAIT QGAHDYLGRY HNQISVIEEG RSKELFGWVA PQPDKYSITR
351  TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV
401  GDTDSAQALG CLELDEEDLA LCSFVCPGKY EXGPLLRKVL ETXEKEG*
```

The originally-identified partial strain B sequence (ORF22) shows 94.2% identity over a 158aa overlap with ORF22a:

```
                    10        20        30        40        50        60
orf22.pep   MIKIKKGLNLPIAGRPEQAVYDGPAITEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED
            ||||||||||||||||||::||||:|||||||||||||||| ||||||||||||||||||
orf22a      MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMKVKEGDAVKKGQVLFED
```

```
                        10        20        30        40        50        60

                        70        80        90       100       110       120
orf22.pep      KKNPGVVFTAPASGKIAAIHRGEKRVLQSVVIAVEXNDEIEFERYAPEALANLSGEEVRR
               || ||||||||:||||||||||||||||||||||| |||||||||||||||||||| |
orf22a         KKXPGVVFTAPVSGKIAAIHRGEKRVLQSVVIAVEGNDEIEFERYAPEALANLSGXEXXX
                        70        80        90       100       110       120

                       130       140       150
orf22.pep      NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP
               ||||||||||||:|||||||||||||||||||||||||
orf22a         NLIQSGLWTALRXRPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRRXXLV
                       130       140       150       160       170       180
```

The complete strain B sequence (ORF22-1) and ORF22a show 94.9% identity in 447 aa overlap:

```
                        10        20        30        40        50        60
orf22a.pep     MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMKVKEGDAVKKGQVLFED
               ||||||||||||||||||::||||:|||||||||||||||| ||||||||||||||||||
orf22-1        MIKIKKGLNLPIAGRPEQAVYDGPAITEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED
                        10        20        30        40        50        60

                        70        80        90       100       110       120
orf22a.pep     KKXPGVVFTAPVSGKIAAIHRGEKRVLQSVVIAVEGNDEIEFERYAPEALANLSGXEXXX
               || ||||||||:|||||||||||||||||||||||||||||||||||||||||||| |
orf22-1        KKNPGVVFTAPASGKIAAIHRGEKRVLQSVVIAVEGNDEIEFERYAPEALANLSGEEVRR
                        70        80        90       100       110       120

                       130       140       150       160       170       180
orf22a.pep     NLIQSGLWTALRXRPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRRXXLV
               ||||||||||||:||||||||||||||||||||||||||||||:|:||||  ||:|  ||
orf22-1        NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV
                       130       140       150       160       170       180

                       190       200       210       220       230       240
orf22a.pep     LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFIEPVGANKTVWTI
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf22-1        LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFIEPVGANKTVWTI
                       190       200       210       220       230       240

                       250       260       270       280       290       300
orf22a.pep     NYQDVIAIGRLFATGRLNTERVIALGGSQVNKPRLLRTVLGAKVSQITAGELVDADNRVI
               ||||||:|||||||||||||||||||||||||||||||||||||||||||||||:|||||
orf22-1        NYQDVITIGRLFATGRLNTERVIALGGSQVNKPRLLRTVLGAKVSQITAGELVDTDNRVI
                       250       260       270       280       290       300

                       310       320       330       340       350       360
orf22a.pep     SGSVLNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPDKYSITRTTLGHFLKNK
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf22-1        SGSVLNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPDKYSITRTTLGHFLKNK
                       310       320       330       340       350       360

                       370       380       390       400       410       420
orf22a.pep     LFKFTTAVNGGDRAMVPIGTYERVMPLDILPTLLLRDLIVGDTDSAQALGCLELDEEDLA
               ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf22-1        LFKFNTAVNGGDRAMVPIGTYERVMPLDILPTLLLRDLIVGDTDSAQALGCLELDEEDLA
                       370       380       390       400       410       420

                       430       440
orf22a.pep     LCSFVCPGKYEXGPLLRKVLETXEKEGX
               ||||||||||| |||||||||| |||||
orf22-1        LCSFVCPGKYEYGPLLRKVLETIEKEGX
                       430       440
```

Further work identified a partial gene sequence <SEQ ID 129> from *N.gonorrhoeae*, which encodes the following amino acid sequence <SEQ ID 130; ORF22ng>:

```
  1  MIKIKKGLNL PIAGRPEQVI YDGPAITEVA LLGEEYVGMR PSMKIKEGEA
 51  VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLQSV VIAVEGNDEI
101  EFERYVPEAL AKLSSEKVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
```

```
151  VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADVP
201  SENAANIETH EFGGPHPAGL SGTHIHFIEP VGANKTVWTI NYQDVIAIGR
251  LFVTGRLNTE RVVALGGLQV NKPRLLRTVL GAKVSQLTAG ELVDADNRVI
301  SGSVLNGAIA QGAHDYLGRY HN*
```

Further work identified complete gonococcal gene <SEQ ID 131>:

```
   1  ATGATTAAAA TCAAAAAAGG TCTAAATCTG CCCATCGCGG GCAGACCGGA
  51  GCAAGTCATT TATGACGGCC CGGCCATTAC CGAAGTCGCG TTGCTTGGCG
 101  AAGAATATGT CGGCATGCGC CCCTCGATGA AAATCAAGGA AGGTGAAGCC
 151  GTCAAAAAAG GCCAAGTGCT GTTTGAAGAC AAAAAGAATC CGGGCGTAGT
 201  ATTTACTGCG CCGGCTTCAG GCAAAATCGC CGCTATTCAC CGTGGCGAAA
 251  AGCGCGTACT TCAGTCAGTC GTGATTGCCG TTGAAGGCAA CGACGAAATC
 301  GAGTTCGAAC GCTACGTACC TGAAGCGCTG GCAAAATTGA GCAGCGAAAA
 351  AGTGCGCCGC AACCTGATTC AATCAGGCTT ATGGACTGCG CTTCGCACCC
 401  GTCCGTTCAG CAAAATCCCT GCCGTAGATG CCGAGCCGTT CGCCATCTTC
 451  GTCAATGCGA TGGACACCAA TCCGCTGGCT GCCGACCCTA CGGTCATCAT
 501  CAAAGAAGCC GCCGAAGACT TCAAACGCGG CCTGTTGGTA TTGAGCCGCC
 551  TGACCGAACG TAAAATCCAT GTGTGTAAAG CAGCAGGCGC AGACGTGCCG
 601  TCTGAAAATG CTGCCAATAT CGAAACACAT GAATTTGGCG GCCCGCATCC
 651  TGCCGGCTTG AGTGGCACGC ACATTCATTT CATCGAGCCA GTCGGCGCGA
 701  ATAAAACCGT GTGGACCATC AATTATCAAG ACGTGATTGC TATCGGACGT
 751  TTGTTCGTAA CAGGCCGTCT GAATACCGAG CGCGTGGTTG CCTTGGGCGG
 801  CCTGCAAGTC AACAAACCGC GCCTCTTGCG TACCGTTTTG GGTGCGAAGG
 851  TGTCTCAACT TACCGCCGGC GAATTGGTTG ACGCGGACAA CCGCGTGATT
 901  TCCGGTTCGG TATTGAACGG TGCGATTGCA CAAGGCGCGC ATGATTATTT
 951  GGGACGCTAC CACAATCAGA TTTCCGTTAT CGAAGAAGGC CGCAGCAAAG
1001  AGCTGTTCGG CTGGGTTGCG CCGCAGCCGG ACAAATACTC CATCACGCGC
1051  ACCACTCTCG GCCATTTCCT AAAAAACAAA CTCTTCAAGT TCACGACAGC
1101  CGTCAACGGC GGCGACCGCG CCATGGTACC GATCGGCACT TATGAGCGCG
1151  TAATGCCGTT GGACATCCTG CCTACCTTGC TTTTGCGCGA TTTAATCGTC
1201  GGCGATACCG ACAGCGCGCA GGCTTTGGGT TGCTTGGAAT TGGACGAAGA
1251  AGACCTCGCT TTGTGCAGCT TCGTCTGCCC GGGCAAATAC GAATACGGCC
1301  CGCTGTTGCG CAAAGTGCTG GAAACCATTG AGAAGGAAGG CTGA
```

This encodes a protein having amino acid sequence <SEQ ID 132; ORF22ng-1>:

```
  1  MIKIKKGLNL PIAGRPEQVI YDGPAITEVA LLGEEYVGMR PSMKIKEGEA
 51  VKKGQVLFED KKNPGVVFTA PASGKIAAIH RGEKRVLQSV VIAVEGNDEI
101  EFERYVPEAL AKLSSEKVRR NLIQSGLWTA LRTRPFSKIP AVDAEPFAIF
151  VNAMDTNPLA ADPTVIIKEA AEDFKRGLLV LSRLTERKIH VCKAAGADVP
201  SENAANIETH EFGGPHPAGL SGTHIHFIEP VGANKTVWTI NYQDVIAIGR
251  LFVTGRLNTE RVVALGGLQV NKPRLLRTVL GAKVSQLTAG ELVDADNRVI
301  SGSVLNGAIA QGAHDYLGRY HNQISVIEEG RSKELFGWVA PQPDKYSITR
351  TTLGHFLKNK LFKFTTAVNG GDRAMVPIGT YERVMPLDIL PTLLLRDLIV
401  GDTDSAQALG CLELDEEDLA LCSFVCPGKY EYGPLLRKVL ETIEKEG*
```

The originally-identified partial strain B sequence (ORF22) shows 93.7% identity over a 158aa overlap with ORF22ng:

```
orf22.pep   MIKIKKGLNLPIAGRPEQAVYDGPAITEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED   60
            ||||||||||||||||||::||||||||||||||||:|||||||:|||:|||||||||||
orf22ng     MIKIKKGLNLPIAGRPEQVIYDGPAITEVALLGEEYVGMRPSMKIKEGEAVKKGQVLFED   60

orf22.pep   KKNPGVVFTAPASGKIAAIHRGEKRVLQSVVIAVEXNDEIEFERYAPEALANLSGEEVRR  120
            ||||||||||||||||||||||||||||||||||| |||||||||:|||||:||:|:|||
orf22ng     KKNPGVVFTAPASGKIAAIHRGEKRVLQSVVIAVEGNDEIEFERYVPEALAKLSSEKVRR  120

orf22.pep   NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP                        158
            ||||||||||||||||||||||||||||||||||||||
orf22ng     NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV  180
```

The complete sequences from strain B (ORF22-1) and gonococcus (ORF22ng) show 96.2% identity in 447 aa overlap:

```
                      10        20        30        40        50        60
orf22-1.pep   MIKIKKGLNLPIAGRPEQAVYDGPAITEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED
```

```
                ||||||||||||||||||::||||||||||||||||:|||||||:|||:|||||||||||
orf22ng-1       MIKIKKGLNLPIAGRPEQVIYDGPAITEVALLGEEYVGMRPSMKIKEGEAVKKGQVLFED
                        10        20        30        40        50        60

                        70        80        90       100       110       120
orf22-1.pep     KKNPGVVFTAPASGKIAAIHRGEKRVLQSVVIAVEGNDEIEFERYAPEALANLSGEEVRR
                ||||||||||||||||||||||||||||||||||||||||||||:||||||:||:|:|||
orf22ng-1       KKNPGVVFTAPASGKIAAIHRGEKRVLQSVVIAVEGNDEIEFERYVPEALAKLSSEKVRR
                        70        80        90       100       110       120

                       130       140       150       160       170       180
orf22-1.pep     NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf22ng-1       NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV
                       130       140       150       160       170       180

                       190       200       210       220       230       240
orf22-1.pep     LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFIEPVGANKTVWTI
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf22ng-1       LSRLTERKIHVCKAAGADVPSENAANIETHEFGGPHPAGLSGTHIHFIEPVGANKTVWTI
                       190       200       210       220       230       240

                       250       260       270       280       290       300
orf22-1.pep     NYQDVITIGRLFATGRLNTERVIALGGSQVNKPRLLRTVLGAKVSQITAGELVDTDNRVI
                ||||||:|||||:|||||||||:|||| ||||||||||||||||||:|||||||:|||||
orf22ng-1       NYQDVIAIGRLFVTGRLNTERVVALGGLQVNKPRLLRTVLGAKVSQLTAGELVDADNRVI
                       250       260       270       280       290       300

                       310       320       330       340       350       360
orf22-1.pep     SGSVLNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPDKYSITRTTLGHFLKNK
                |||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||
orf22ng-1       SGSVLNGAIAQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPDKYSITRTTLGHFLKNK
                       310       320       330       340       350       360

                       370       380       390       400       410       420
orf22-1.pep     LFKFNTAVNGGDRAMVPIGTYERVMPLDILPTLLLRDLIVGDTDSAQALGCLELDEEDLA
                ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf22ng-1       LFKFTTAVNGGDRAMVPIGTYERVMPLDILPTLLLRDLIVGDTDSAQALGCLELDEEDLA
                       370       380       390       400       410       420

                       430       440
orf22-1.pep     LCSFVCPGKYEYGPLLRKVLETIEKEGX
                |||||||||||||||||||||||||||
orf22ng-1       LCSFVCPGKYEYGPLLRKVLETIEKEGX
                       430       440
```

Computer analysis of these sequences gave the following results:

Homology with 48kDa outer membrane protein of *Actinobacillus pleuropneumoniae* (accession number U24492).

ORF22 and this 48kDa protein show 72% aa identity in 158aa overlap:

```
Orf22  1    MIKIKKGLNLPIAGRPEQAVYDGPAITEVALLGEEYAGMRPSMKVKEGDAVKKGQVLFED 60
            MI IKKGL+LPIAG P Q +++G  + EVA+LGEEY GMRPSMKV+EGD VKKGQVLFED
48kDa  1    MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMRPSMKVREGDVVKKGQVLFED 60

orf22  61   KKNPGVVFTAPASGKIAAIHRGEKRVLQSVVIAVEXNDEIEFERYAPEALANLSGEEVRR 120
            KKNPGVVFTAPASG +  I+RGEKRVLQSVVI VE +++I F RY    LA+LS E+V++
48kDa  61   KKNPGVVFTAPASGTVVTINRGEKRVLQSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120

orf22  121 NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNP 158
           NLI+SGLWTA RTRPFSK+PA+DA P +IFVNAMDTNP
48kDa  121 NLIESGLWTAFRTRPFSKVPALDAIPSSIFVNAMDTNP 158
```

ORF22a also shows homology to the 48kDa *Actinobacillus pleuropneumoniae* protein:

```
gi|1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus pleuropneumoniae]
Length = 449

 Score =  530 bits (1351), Expect = e-150
```

```
Identities = 274/450 (60%), Positives = 323/450 (70%), Gaps = 4/450 (0%)

Query: 1   MIKIKKGLNLPIAGRPEQVIYDGPVITEVALLGEEYAGMRPXMKVKEGDAVKKGQVLFED 60
            MI IKKGL+LPIAG P QVI++G  + EVA+LGEEY GMRP MKV+EGD VKKGQVLFED
Sbjct: 1   MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMRPSMKVREGDVVKKGQVLFED 60

Query: 61  KKXPGVVFTAPVSGKIAAIHRGEKRVLQSVVIAVEGNDEIEFERYAPEALANLSGXEXXX 120
            KK PGVVFTAP SG +  I+RGEKRVLQSVVI VEG+++I F RY    LA+LS  +
Sbjct: 61  KKNPGVVFTAPASGTVVTINRGEKRVLQSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120

Query: 121 NLIQSGLWTALRXRPFSKIPAVDAEPFAIFVNAMDTNPLAADPVVVIKEAXXDFRRXXLV 180
            NLI+SGLWTA R RPFSK+PA+DA P +IFVNAMDTNPLAADP VV+KE   DF+   V
Sbjct: 121 NLIESGLWTAFRTRPFSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGLTV 180

Query: 181 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHPAGLSGTHIHFIEPVGANKTV 237
            L+RL    ++ +++CK A +++P S      I    F G HPAGL GTHIHF++PVGA K V
Sbjct: 181 LTRLFNGQKPVYLCKDADSNIPLSPAIEGITIKSFSGVHPAGLVGTHIHFVDPVGATKQV 240

Query: 238 WTINYQDVIAIGRLFATGRLNTERVIALGGSQVNKPRLLRTVLGAKVSQITAGELVDADN 297
            W +NYQDVIAIG+LF TG L T+R+I+L G QV  PRL+RT LGA +SQ+TA EL   +N
Sbjct: 241 WHLNYQDVIAIGKLFTTGELFTDRIISLAGPQVKNPRLVRTRLGANLSQLTANELNAGEN 300

Query: 298 RVISGSVLNGAITQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPDKYSITRTTLGHFL 357
            RVISGSVL+GA   G  DYLGRY  Q+SV+ EGR KELFGW+ P  DK+SITRT LGHF
Sbjct: 301 RVISGSVLSGATAAGPVDYLGRYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360

Query: 358 KNKLFKFTTAVNGGDRAMVPIGTYERVMXXXXXXXXXXXXXXVGDTDSAQXXXXXXXXXX 417
            K KLF FTTAV+GG+RAMVPIG YERVM              GDTDSAQ
Sbjct: 361 K-KLFNFTTAVHGGERAMVPIGAYERVMPLDIIPTLLLRDLAAGDTDSAQNLGCLELDEE 419

Query: 418 XXXXXSFVCPGKYEXGPLLRKVLETXEKEG 447
                 ++VCPGK   GP+LR  LE  EKEG
```

ORF22ng-1 also shows homology with the OMP from *A.pleuropneumoniae*:

```
gi|1185395 (U24492) 48 kDa outer membrane protein [Actinobacillus
pleuropneumoniae] Length = 449
 Score =  555 bits (1414), Expect = e-157
 Identities = 284/450 (63%), Positives = 337/450 (74%), Gaps = 4/450 (0%)

Query: 27  MIKIKKGLNLPIAGRPEQVIYDGPAITEVALLGEEYVGMRPSMKIKEGEAVKKGQVLFED 86
            MI IKKGL+LPIAG P QVI++G  + EVA+LGEEYVGMRPSMK++EG+ VKKGQVLFED
Sbjct: 1   MITIKKGLDLPIAGTPAQVIHNGNTVNEVAMLGEEYVGMRPSMKVREGDVVKKGQVLFED 60

Query: 87  KKNPGVVFTAPASGKIAAIHRGEKRVLQSVVIAVEGNDEIEFERYVPEALAKLSSEKVRR 146
            KKNPGVVFTAPASG +  I+RGEKRVLQSVVI VEG+++I F RY    LA LS+E+V++
Sbjct: 61  KKNPGVVFTAPASGTVVTINRGEKRVLQSVVIKVEGDEQITFTRYEAAQLASLSAEQVKQ 120

Query: 147 NLIQSGLWTALRTRPFSKIPAVDAEPFAIFVNAMDTNPLAADPTVIIKEAAEDFKRGLLV 206
            NLI+SGLWTA RTRPFSK+PA+DA P +IFVNAMDTNPLAADP V++KE   DFK GL V
Sbjct: 121 NLIESGLWTAFRTRPFSKVPALDAIPSSIFVNAMDTNPLAADPEVVLKEYETDFKDGLTV 180

Query: 207 LSRL--TERKIHVCKAAGADVP-SENAANIETHEFGGPHPAGLSGTHIHFIEPVGANKTV 263
            L+RL    ++ +++CK A +++P S      I    F G HPAGL GTHIHF++PVGA K V
Sbjct: 181 LTRLFNGQKPVYLCKDADSNIPLSPAIEGITIKSFSGVHPAGLVGTHIHFVDPVGATKQV 240

Query: 264 WTINYQDVIAIGRLFVTGRLNTERVVALGGLQVNKPRLLRTVLGAKVSQLTAGELVDADN 323
            W +NYQDVIAIG+LF TG L T+R+++L G QV  PRL+RT LGA +SQLTA EL   +N
Sbjct: 241 WHLNYQDVIAIGKLFTTGELFTDRIISLAGPQVKNPRLVRTRLGANLSQLTANELNAGEN 300

Query: 324 RVISGSVLNGAIAQGAHDYLGRYHNQISVIEEGRSKELFGWVAPQPDKYSITRTTLGHFL 383
            RVISGSVL+GA A G  DYLGRY  Q+SV+ EGR KELFGW+ P  DK+SITRT LGHF
Sbjct: 301 RVISGSVLSGATAAGPVDYLGRYALQVSVLAEGREKELFGWIMPGSDKFSITRTVLGHFG 360

Query: 384 KNKLFKFTTAVNGGDRAMVPIGTYERVMXXXXXXXXXXXXXXVGDTDSAQXXXXXXXXXX 443
            K KLF FTTAV+GG+RAMVPIG YERVM              GDTDSAQ
Sbjct: 361 K-KLFNFTTAVHGGERAMVPIGAYERVMPLDIIPTLLLRDLAAGDTDSAQNLGCLELDEE 419

Query: 444 XXXXXSFVCPGKYEYGPLLRKVLETIEKEG 473
                 ++VCPGK  YGP+LR  LE IEKEG
Sbjct: 420 DLALCTYVCPGKNNYGPMLRAALEKIEKEG 449
```

Based on this analysis, including the homology with the outer membrane protein of *Actinobacillus pleuropneumoniae*, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF22-1 (35.4kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 5A shows the results of affinity purification of the GST-fusion protein, and Figure 5B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result) and FACS analysis (Figure 5C). These experiments confirm that ORF22-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 16

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 133>:

```
  1  ..GCGnCGnAAA TCATCCATCC CC..nACGTC GTAGGCCCTG AAGCCAACTG
 51    GTTTTTTATG GTAGCCAGTA CGTTTGTGAT TGCTTTGATT GGTTATTTTG
101    TTACTGAAAA AATCGTCGAA CCGCAATTGG GCCCTTATCA ATCAGATTTG
151    TCACAAGAAG AAAAAGACAT TCGGCATTCC AATGAAATCA CGCCTTTGGA
201    ATATAAAGGA TTAATTTGGG CTGGCGTGGT GTTTGTTGCC TTATCCGCCC
251    TATTGGCTTG GAGCATCGTC CCTGCCGACG GTATTTTGCG TCATCCTGAA
301    ACAGGATTGG TTTCCGGTTC GCCGTTTTTA AAATCGATTG TTGTTTTTAT
351    TTTCTTGTTG TTTGCACTGC CGGGCATTGT TTATGGCCGG GTAACCCGAA
401    GTTTGCGCGG CGAACAGGAA GTCGTTAATG CGmyGGCCGA ATCGATGAGT
451    ACTCTGGsGC TTTmTTTGsw CAkcATCTTT TTTGCCGCAC AGTTTGTCGC
501    ATTTTTTAAT TGGACGAATA TTGGGCAATA TATTGCCGTT AAAGGGGCGA
551    CGTTCTTAAA AGAAGTCGGC TTGGGCGGCA GCGTGTTGTT TATCGGTTTT
601    ATTTTAATTT GTGCTTTTAT CAATCTGATG ATAGGCTCCG CCTCCGCGCA
651    ATGGGCGGTA ACTGCGCCGA TTTTCGTCCC TATGCTGATG TTGGCCGGCT
701    ACGCGCCCGA AGTCATTCAA GCCGCTTACC GCATCGGTGA TTCCGTTACC
751    AATATTATTA CGCCGATGAT GAGTTATTTC GGGCTGATTA TGGCGACGGT
801    GrkCmmmTAC AAAAAAGATG CGGGCGTGGG TaCGcTGATT wCTATGATGT
851    TGCCGTATTC CGCTTTCTTC TTGATTGCgT GGATTGCCTT ATTCTGCATT
901    TGGGTATTTg TTTTGGGCCT GCCCGTCGGT CCCGGCGCGC CCACATTCTA
951    TCCCGCACCT TAA
```

This corresponds to the amino acid sequence <SEQ ID 134; ORF12>:

```
  1  ..AXXIIHPXXV VGPEANWFFM VASTFVIALI GYFVTEKIVE PQLGPYQSDL
 51    SQEEKDIRHS NEITPLEYKG LIWAGVVFVA LSALLAWSIV PADGILRHPE
101    TGLVSGSPFL KSIVVFIFLL FALPGIVYGR VTRSLRGEQE VVNAXAESMS
151    TLXLXLXXIF FAAQFVAFFN WTNIGQYIAV KGATFLKEVG LGGSVLFIGF
201    ILICAFINLM IGSASAQWAV TAPIFVPMLM LAGYAPEVIQ AAYRIGDSVT
251    NIITPMMSYF GLIMATVXXY KKDAGVGTLI XMMLPYSAFF LIAWIALFCI
301    WVFVLGLPVG PGAPTFYPAP *
```

Further sequence analysis revealed the complete DNA sequence <SEQ ID 135> to be:

```
  1  ATGAGTCAAA CCGATACGCA ACGGGACGGA CGATTTTTAC GCACAGTCGA
 51  ATGGCTGGGC AATATGTTGC CGCATCCGGT TACGCTTTTT ATTATTTTCA
101  TTGTGTTATT GCTGATTGCC TCTGCCGTCG GTGCGTATTT CGGACTATCC
151  GTCCCCGATC CGCGCCCTGT TGGTGCGAAA GGACGTGCCG ATGACGGTTT
201  GATTTACATT GTCAGCCTGC TCAATGCCGA CGGTTTTATC AAAATCCTGA
251  CGCATACCGT TAAAAATTTC ACCGGTTTCG CGCCGTTGGG AACGGTGTTG
301  GTTTCTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTTCCGC
351  ATTAATGCGC TTATTGCTCA CAAAATCGCC ACGCAAACTC ACTACTTTTA
401  TGGTTGTTTT TACAGGGATT TTATCTAATA CCGCTTCTGA ATTGGGCTAT
451  GTCGTCCTAA TCCCTTTGTC CGCCATCATC TTTCATTCCC TCGGCCGCCA
501  TCCGCTTGCC GGTCTGGCTG CGGCTTTCGC CGGCGTTTCG GGCGGTTATT
```

```
 551  CGGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTTGGC AGGCATCACC
 601  CAACAGGCGG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC
 651  CAACTGGTTT TTTATGGTAG CCAGTACGTT TGTGATTGCT TTGATTGGTT
 701  ATTTTGTTAC TGAAAAAATC GTCGAACCGC AATTGGGCCC TTATCAATCA
 751  GATTTGTCAC AAGAAGAAAA AGACATTCGG CATTCCAATG AAATCACGCC
 801  TTTGGAATAT AAAGGATTAA TTTGGGCTGG CGTGGTGTTT GTTGCCTTAT
 851  CCGCCCTATT GGCTTGGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
 901  CCTGAAACAG GATTGGTTTC CGGTTCGCCG TTTTTAAAAT CGATTGTTGT
 951  TTTTATTTTC TTGTTGTTTG CACTGCCGGG CATTGTTTAT GGCCGGGTAA
1001  CCCGAAGTTT GCGCGGCGAA CAGGAAGTCG TTAATGCGAT GGCCGAATCG
1051  ATGAGTACTC TGGGGCTTTA TTTGGTCATC ATCTTTTTTG CCGCACAGTT
1101  TGTCGCATTT TTTAATTGGA CGAATATTGG GCAATATATT GCCGTTAAAG
1151  GGGCGACGTT CTTAAAAGAA GTCGGCTTGG GCGGCAGCGT GTTGTTTATC
1201  GGTTTTATTT TAATTTGTGC TTTTATCAAT CTGATGATAG GCTCCGCCTC
1251  CGCGCAATGG GCGGTAACTG CGCCGATTTT CGTCCCTATG CTGATGTTGG
1301  CCGGCTACGC GCCCGAAGTC ATTCAAGCCG CTTACCGCAT CGGTGATTCC
1351  GTTACCAATA TTATTACGCC GATGATGAGT TATTTCGGGC TGATTATGGC
1401  GACGGTGATC AAATACAAAA AAGATGCGGG CGTGGGTACG CTGATTTCTA
1451  TGATGTTGCC GTATTCCGCT TTCTTCTTGA TTGCGTGGAT TGCCTTATTC
1501  TGCATTTGGG TATTTGTTTT GGGCCTGCCC GTCGGTCCCG GCGCGCCCAC
1551  ATTCTATCCC GCACCTTAA
```

This corresponds to the amino acid sequence <SEQ ID 136; ORF12-1>:

```
  1  MSQTDTQRDG RFLRTVEWLG NMLPHPVTLF IIFIVLLLIA SAVGAYFGLS
 51  VPDPRPVGAK GRADDGLIYI VSLLNADGFI KILTHTVKNF TGFAPLGTVL
101  VSLLGVGIAE KSGLISALMR LLLTKSPRKL TTFMVVFTGI LSNTASELGY
151  VVLIPLSAII FHSLGRHPLA GLAAAFAGVS GGYSANLFLG TIDPLLAGIT
201  QQAAQIIHPD YVVGPEANWF FMVASTFVIA LIGYFVTEKI VEPQLGPYQS
251  DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
301  PETGLVSGSP FLKSIVVFIF LLFALPGIVY GRVTRSLRGE QEVVNAMAES
351  MSTLGLYLVI IFFAAQFVAF FNWTNIGQYI AVKGATFLKE VGLGGSVLFI
401  GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGYAPEV IQAAYRIGDS
451  VTNIITPMMS YFGLIMATVI KYKKDAGVGT LISMMLPYSA FFLIAWIALF
501  CIWVFVLGLP VGPGAPTFYP AP*
```

Computer analysis of this amino acid sequence gave the following results:

<u>Homology with a predicted ORF from *N.meningitidis* (strain A)</u>

ORF12 shows 96.3% identity over a 320aa overlap with an ORF (ORF12a) from strain A of *N. meningitidis:*

```
                                                     10        20        30
orf12.pep                               AXXIIHPXXVVGPEANWFFMVASTFVIALI
                                        |  ||||  ||||||||||||||||||||||
orf12a     AAAFAGVSGGYSANLFLGTIDPLLAGITQQAAQIIHPDYVVGPEANWFFMVASTFVIALI
              180       190       200       210       220       230

                   40        50        60        70        80        90
orf12.pep  GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf12a     GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV
              240       250       260       270       280       290

                  100       110       120       130       140       150
orf12.pep  PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAXAESMS
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
orf12a     PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAMAESMS
              300       310       320       330       340       350

                  160       170       180       190       200       210
orf12.pep  TLXLXLXXIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLM
           || | |  ||||||||||||||||||||||||||||||||||||||||||||||||||||
orf12a     TLGLYLVIIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLM
              360       370       380       390       400       410

                  220       230       240       250       260       270
orf12.pep  IGSASAQWAVTAPIFVPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVXXY
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |
```

```
orf12a     IGSASAQWAVTAPIFVPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKY
                   420       430       440       450       460       470

                   280       290       300       310       320
orf12.pep  KKDAGVGTLIXMMLPYSAFFLIAWIALFCIWVFVLGLPVGPGAPTFYPAPX
           ||||||||| |||||||||||||||||||||||||||||||||||||||||
orf12a     KKDAGVGTLISMMLPYSAFFLIAWIALFCIWVFVLGLPVGPGAPTFYPAPX
                   480       490       500       510       520
```

The complete length ORF12a nucleotide sequence <SEQ ID 137> is:

```
   1  ATGAGTCAAA CCGATACGCA ACGGGACGGA CGATTTTTAC GCACAGTCGA
  51  ATGGCTGGGC AATATGTTGC CGCACCCGGT TACGCTTTTT ATTATTTTCA
 101  TTGTGTTATT GCTGATTGCC TCTGCCGCCG GTGCGTATTT CGGACTATCC
 151  GTCCCCGATC CGCGCCCTGT TGGTGCGAAA GGACGTGCCG ATGACGGTTT
 201  GATTCACGTT GTCAGCCTGC TCGATGCTGA CGGTTTGATC AAAATCCTGA
 251  CGCATACCGT TAAAAATTTC ACCGGTTTCG CGCCGTTGGG AACGGTGTTG
 301  GTTTCTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTTCCGC
 351  ATTAATGCGC TTATTGCTCA CAAAATCTCC ACGCAAACTC ACTACTTTTA
 401  TGGTTGTTTT TACAGGGATT TTATCTAATA CCGCTTCTGA ATTGGGCTAT
 451  GTCGTCCTAA TCCCTTTGTC CGCCATCATC TTTCATTCCC TCGGCCGCCA
 501  TCCGCTTGCC GGTCTGGCTG CGGCTTTCGC CGGCGTTTCG GGCGGTTATT
 551  CGGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTTGGC AGGCATCACC
 601  CAACAGGCGG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC
 651  CAACTGGTTT TTTATGGTAG CCAGTACGTT TGTGATTGCT TTGATTGGTT
 701  ATTTTGTTAC TGAAAAAATC GTCGAACCGC AATTGGGCCC TTATCAATCA
 751  GATTTGTCAC AAGAAGAAAA AGACATTCGA CATTCCAATG AAATCACGCC
 801  TTTGGAATAT AAAGGATTAA TTTGGGCTGG CGTGGTGTTT GTTGCCTTAT
 851  CCGCCCTATT GGCTTGGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
 901  CCTGAAACAG GATTGGTTTC CGGTTCGCCG TTTTTAAAAT CAATTGTTGT
 951  TTTTATTTTC TTGTTGTTTG CACTGCCGGG CATTGTTTAT GGCCGGGTAA
1001  CCCGAAGTTT GCGCGGCGAA CAGGAAGTCG TTAATGCGAT GGCCGAATCG
1051  ATGAGTACTC TGGGGCTTTA TTTGGTCATC ATCTTTTTTG CCGCACAGTT
1101  TGTCGCATTT TTTAATTGGA CGAATATTGG GCAATATATT GCCGTTAAAG
1151  GGGCGACGTT CTTAAAAGAA GTCGGCTTGG GCGGCAGCGT GTTGTTTATC
1201  GGTTTTATTT TAATTTGTGC TTTTATCAAT CTGATGATAG GCTCCGCCTC
1251  CGCGCAATGG GCGGTAACTG CGCCGATTTT CGTCCCTATG CTGATGTTGG
1301  CCGGCTACGC GCCCGAAGTC ATTCAAGCCG CTTACCGCAT CGGTGATTCC
1351  GTTACCAATA TTATTACGCC GATGATGAGT TATTTCGGGC TGATTATGGC
1401  GACGGTGATC AAATACAAAA AAGATGCGGG CGTGGGTACG CTGATTTCTA
1451  TGATGTTGCC GTATTCCGCT TTCTTCTTGA TTGCGTGGAT TGCCTTATTC
1501  TGCATTTGGG TATTTGTTTT GGGCCTGCCC GTCGGTCCCG GCGCGCCCAC
1551  ATTCTATCCC GCACCTTAA
```

This encodes a protein having amino acid sequence <SEQ ID 138>:

```
  1  MSQTDTQRDG RFLRTVEWLG NMLPHPVTLF IIFIVLLLIA SAAGAYFGLS
 51  VPDPRPVGAK GRADDGLIHV VSLLDADGLI KILTHTVKNF TGFAPLGTVL
101  VSLLGVGIAE KSGLISALMR LLLTKSPRKL TTFMVVFTGI LSNTASELGY
151  VVLIPLSAII FHSLGRHPLA GLAAAFAGVS GGYSANLFLG TIDPLLAGIT
201  QQAAQIIHPD YVVGPEANWF FMVASTFVIA LIGYFVTEKI VEPQLGPYQS
251  DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
301  PETGLVSGSP FLKSIVVFIF LLFALPGIVY GRVTRSLRGE QEVVNAMAES
351  MSTLGLYLVI IFFAAQFVAF FNWTNIGQYI AVKGATFLKE VGLGGSVLFI
401  GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGYAPEV IQAAYRIGDS
451  VTNIITPMMS YFGLIMATVI KYKKDAGVGT LISMMLPYSA FFLIAWIALF
501  CIWVFVLGLP VGPGAPTFYP AP*
```

ORF12a and ORF12-1 show 99.0% identity in 522 aa overlap:

```
                    10        20        30        40        50        60
orf12a.pep  MSQTDTQRDGRFLRTVEWLGNMLPHPVTLFIIFIVLLLIASAAGAYFGLSVPDPRPVGAK
            ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
orf12-1     MSQTDTQRDGRFLRTVEWLGNMLPHPVTLFIIFIVLLLIASAVGAYFGLSVPDPRPVGAK
                    10        20        30        40        50        60

                    70        80        90       100       110       120
orf12a.pep  GRADDGLIHVVSLLDADGLIKILTHTVKNFTGFAPLGTVLVSLLGVGIAEKSGLISALMR
            ||||||||::||||:|||:|||||||||||||||||||||||||||||||||||||||||
orf12-1     GRADDGLIYIVSLLNADGFIKILTHTVKNFTGFAPLGTVLVSLLGVGIAEKSGLISALMR
```

```
                        70        80        90       100       110       120

                       130       140       150       160       170       180
orf12a.pep      LLLTKSPRKLTTFMVVFTGILSNTASELGYVVLIPLSAIIFHSLGRHPLAGLAAAFAGVS
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf12-1         LLLTKSPRKLTTFMVVFTGILSNTASELGYVVLIPLSAIIFHSLGRHPLAGLAAAFAGVS
                       130       140       150       160       170       180

                       190       200       210       220       230       240
orf12a.pep      GGYSANLFLGTIDPLLAGITQQAAQIIHPDYVVGPEANWFFMVASTFVIALIGYFVTEKI
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf12-1         GGYSANLFLGTIDPLLAGITQQAAQIIHPDYVVGPEANWFFMVASTFVIALIGYFVTEKI
                       190       200       210       220       230       240

                       250       260       270       280       290       300
orf12a.pep      VEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf12-1         VEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH
                       250       260       270       280       290       300

                       310       320       330       340       350       360
orf12a.pep      PETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAMAESMSTLGLYLVI
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf12-1         PETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAMAESMSTLGLYLVI
                       310       320       330       340       350       360

                       370       380       390       400       410       420
orf12a.pep      IFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLMIGSASAQW
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf12-1         IFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLMIGSASAQW
                       370       380       390       400       410       420

                       430       440       450       460       470       480
orf12a.pep      AVTAPIFVPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGT
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf12-1         AVTAPIFVPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGT
                       430       440       450       460       470       480

                       490       500       510       520
orf12a.pep      LISMMLPYSAFFLIAWIALFCIWVFVLGLPVGPGAPTFYPAPX
                |||||||||||||||||||||||||||||||||||||||||||
orf12-1         LISMMLPYSAFFLIAWIALFCIWVFVLGLPVGPGAPTFYPAPX
                       490       500       510       520
```

<u>Homology with a predicted ORF from *N.gonorrhoeae*</u>

ORF12 shows 92.5% identity over a 320aa overlap with a predicted ORF (ORF12.ng) from *N. gonorrhoeae:*

```
orf12.pep                                     AXXIIHPXXVVGPEANWFFMVASTFVIALI   30
                                              |  ||||  |||||||||||:|||||||||
orf12ng         AAAFAGVSGGYSANLFLGTIDPLLAGITQQAAQIIHPDYVVGPEANWFFMAASTFVIALI  232

orf12.pep       GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV   90
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf12ng         GYFVTEKIVEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIV  292

orf12.pep       PADGILRHPETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAXAESMS  150
                ||||||||||||||:|||||||||||||||||||||||||:|||||||:||||| |||||
orf12ng         PADGILRHPETGLVAGSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNAMAESMS  352

orf12.pep       TLXLXLXXIFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLM  210
                || | |  |||||||||||||||||||||||||:|||:  ||||||||||||||||||||
orf12ng         TLGLYLVIIFFAAQFVAFFNWTNIGQYIAVKGAVFLKKFRLGGSVLFIGFILICAFINLM  412

orf12.pep       IGSASAQWAVTAPIFVPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVXXY  270
                ||||||||||||||||||||||| ||:||||||||||||||||||||||||||||||  |
orf12ng         IGSASAQWAVTAPIFVPMLMLAGNAPQVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKY  472
```

```
orf12.pep    KKDAGVGTLIXMMLPYSAFFLIAWIALFCIWVFVLGLPVGPGAPTFYPAP   320
             |||||||||| |||||||||||||||||||||||||||||||:|||||:|
orf12ng      KKDAGVGTLISMMLPYSAFFLIAWIALFCIWVFVLGLPVGPGTPTFYPVP   522
```

The complete length ORF12ng nucleotide sequence <SEQ ID 139> is:

```
   1  ATGAGTCAAA CCGACGCGCG TCGTAGCGGA CGATTTTTAC GCACAGTCGA
  51  ATGGCTGGGC AATATGTTGC CGCACCCGGT TACGCTTTTT ATTATTTTCA
 101  TTGTGTTATT GCTGATTGcc tctgCCGTCG GTGCGTATTT CGGACTATCC
 151  GTCCCCGATC CGCGTCCTGT TGGGGCGAAA GGACGTGCCG ATGACGGTTT
 201  GATTCACGTT GTCAGCCTGC TCGATGCCGA CGGTTTGATC AAAATCCTGA
 251  CGCATACCGT TAAAAATTTC ACCGGTTTCG CGCCGTTGGG AACGGTGTTG
 301  GTTTCTTTAT TGGGCGTGGG GATTGCGGAA AAATCGGGCT TGATTTCCGC
 351  ATTAATGCGC TTATTGCTCA CAAAATCCCC ACGCAAACTC ACTACTTTTA
 401  TGGTTGTTTT TACAGGGATT TTATCCAATA CGGCTTCTGA ATTGGGCTAT
 451  GTCGTCCTAA TCCCTTTGTC CGCCGTCATC TTTCATTCGC TCGGCCGCCA
 501  TCCGCTTGCC GGTTTGGCTG CGGCTTTCGC CGGCGTTTCG GGCGGTTATT
 551  CGGCCAATCT GTTCTTAGGC ACAATCGATC CGCTCTTGGC AGGCATCACC
 601  CAACAGGCGG CGCAAATCAT CCATCCCGAC TACGTCGTAG GCCCTGAAGC
 651  CAACTGGTTT TTTATGGCAG CCAGTACGTT TGTGATTGCT TTGATTGGTT
 701  ATTTTGTTAC TGAAAAAATC GTCGAACCGC AATTGGGCCC TTATCAATCA
 751  GATTTGTCAC AAGAAGAAAA AGACATTCGG CATTCCAATG AAATCACGCC
 801  TTTGGAATAT AAAGGATTAA TTTGGGCAGG CGTGGTGTTT GTTGCCTTAT
 851  CCGCCCTATT GGCTTGGAGC ATCGTCCCTG CCGACGGTAT TTTGCGTCAT
 901  CCTGAAACAG GATTGGTTGC CGGTTCGCCG TTTTTAAAAT CGATTGTTGT
 951  TTTTATTTTC TTGTTGTTTG CGCTGCCGGG CATTGTTTAT GGCCGGATAA
1001  CCCGAAGTTT GCGCGGCGAA CGGGAAGTCG TTAATGCGAT GGCCGAATCG
1051  ATGAGTACTT TGGGACTTTA TTTGGTCATC ATCTTTTTTG CCGCACAGTT
1101  TGTCGCATTT TTTAATTGGA CGAATATTGG GCAATATATT GCCGTTAAAG
1151  GGGCGGTGTT CTTAAAAGAA GTCGGCTTGG GCGGCAGTGT GTTGTTTATC
1201  GGTTTTATTT TAATTTGTGC TTTTATCAAT CTGATGATAG GCTCCGCCTC
1251  CGCGCAATGG GCGGTAACTG CGCCGATTTT CGTCCCTATG CTGATGTTGG
1301  CCGGCTACGC GCCCGAAGTC ATTCAAGCCG CTTACCGCAT CGGTGATTCC
1351  GTTACCAATA TTATTACGCC GATGATGAGT TATTTCGGGC TGATTATGGC
1401  GACGGTAATC AAATACAAAA AAGATGCGGG CGTAGGCACG CTGATTTCTA
1451  TGATGTTGCC GTATTCCGCT TTCTTCTTAA TTGCATGGAT CGCCTTATTC
1501  TGCATTTGGG TATTTGTTTT GGGTCTGCCC GTCGGTCCCG GCACACCCAC
1551  ATTCTATCCG GTGCCTTAA
```

This encodes a protein having amino acid sequence <SEQ ID 140>:

```
  1  MSQTDARRSG RFLRTVEWLG NMLPHPVTLF IIFIVLLLIA SAVGAYFGLS
 51  VPDPRPVGAK GRADDGLIHV VSLLDADGLI KILTHTVKNF TGFAPLGTVL
101  VSLLGVGIAE KSGLISALMR LLLTKSPRKL TTFMVVFTGI LSNTASELGY
151  VVLIPLSAVI FHSLGRHPLA GLAAAFAGVS GGYSANLFLG TIDPLLAGIT
201  QQAAQIIHPD YVVGPEANWF FMAASTFVIA LIGYFVTEKI VEPQLGPYQS
251  DLSQEEKDIR HSNEITPLEY KGLIWAGVVF VALSALLAWS IVPADGILRH
301  PETGLVAGSP FLKSIVVFIF LLFALPGIVY GRITRSLRGE REVVNAMAES
351  MSTLGLYLVI IFFAAQFVAF FNWTNIGQYI AVKGAVFLKK FRLGGSVLFI
401  GFILICAFIN LMIGSASAQW AVTAPIFVPM LMLAGNAPQV IQAAYRIGDS
451  VTNIITPMMS YFGLIMATVI KYKKDAGVGT LISMMLPYSA FFLIAWIALF
501  CIWVFVLGLP VGPGTPTFYP VP*
```

ORF12ng shows 97.1% identity in 522 aa overlap with ORF12-1:

```
                     10        20        30        40        50        60
orf12-1.pep  MSQTDTQRDGRFLRTVEWLGNMLPHPVTLFIIFIVLLLIASAVGAYFGLSVPDPRPVGAK
             |||||::|:|||||||||||||||||||||||||||||||||||||||||||||||||||
orf12ng      MSQTDARRSGRFLRTVEWLGNMLPHPVTLFIIFIVLLLIASAVGAYFGLSVPDPRPVGAK
                     10        20        30        40        50        60

                     70        80        90       100       110       120
orf12-1.pep  GRADDGLIYIVSLLNADGFIKILTHTVKNFTGFAPLGTVLVSLLGVGIAEKSGLISALMR
             ||||||||::||||:|||:|||||||||||||||||||||||||||||||||||||||||
orf12ng      GRADDGLIHVVSLLDADGLIKILTHTVKNFTGFAPLGTVLVSLLGVGIAEKSGLISALMR
                     70        80        90       100       110       120

                    130       140       150       160       170       180
orf12-1.pep  LLLTKSPRKLTTFMVVFTGILSNTASELGYVVLIPLSAIIFHSLGRHPLAGLAAAFAGVS
             ||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
orf12ng      LLLTKSPRKLTTFMVVFTGILSNTASELGYVVLIPLSAVIFHSLGRHPLAGLAAAFAGVS
```

```
                      130       140       150       160       170       180

                      190       200       210       220       230       240
orf12-1.pep    GGYSANLFLGTIDPLLAGITQQAAQIIHPDYVVGPEANWFFMVASTFVIALIGYFVTEKI
               ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
orf12ng        GGYSANLFLGTIDPLLAGITQQAAQIIHPDYVVGPEANWFFMAASTFVIALIGYFVTEKI
                      190       200       210       220       230       240

                      250       260       270       280       290       300
orf12-1.pep    VEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf12ng        VEPQLGPYQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRH
                      250       260       270       280       290       300

                      310       320       330       340       350       360
orf12-1.pep    PETGLVSGSPFLKSIVVFIFLLFALPGIVYGRVTRSLRGEQEVVNAMAESMSTLGLYLVI
               ||||||:|||||||||||||||||||||||||:|||||||:|||||||||||||||||||
orf12ng        PETGLVAGSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNAMAESMSTLGLYLVI
                      310       320       330       340       350       360

                      370       380       390       400       410       420
orf12-1.pep    IFFAAQFVAFFNWTNIGQYIAVKGATFLKEVGLGGSVLFIGFILICAFINLMIGSASAQW
               |||||||||||||||||||||||||:||||||||||||||||||||||||||||||||||
orf12ng        IFFAAQFVAFFNWTNIGQYIAVKGAVFLKEVGLGGSVLFIGFILICAFINLMIGSASAQW
                      370       380       390       400       410       420

                      430       440       450       460       470       480
orf12-1.pep    AVTAPIFVPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGT
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf12ng        AVTAPIFVPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGT
                      430       440       450       460       470       480

                      490       500       510       520
orf12-1.pep    LISMMLPYSAFFLIAWIALFCIWVFVLGLPVGPGAPTFYPAPX
               ||||||||||||||||||||||||||||||||||:|||||:||
orf12ng        LISMMLPYSAFFLIAWIALFCIWVFVLGLPVGPGTPTFYPVPX
                      490       500       510       520
```

In addition, ORF12ng shows significant homology with a hypotehtical protein from *E.coli*:

```
sp|P46133|YDAH_ECOLI HYPOTHETICAL 55.1 KD PROTEIN IN OGT-DBPA INTERGENIC REGION
>gi|1787597 (AE000231) hypothetical protein in ogt 5'region [Escherichia coli]
Length = 510
 Score =  329 bits (835), Expect = 2e-89
 Identities = 178/507 (35%), Positives = 281/507 (55%), Gaps = 15/507 (2%)

Query: 8   RSGRFLRTVEWLGNMLPHPVTXXXXXXXXXXXASAVGAYFGLSVPDPRPVGAKGRADDGL 67
           +SG+    VE +GN +PHP              +A+ + FG+S  +P         D
Sbjct: 13  QSGKLYGWVERIGNKVPHPFLLFIYLIIVLMVTTAILSAFGVSAKNP--------TDGTP 64

Query: 68  IHVVSLLDADGLIKILTHTVKNFTGFAPXXXXXXXXXXXXIAEKSGLISALMRLLLTKSP 127
           + V +LL  +GL   L + +KNF+GFAP            +AE+ GL+ ALM  + +
Sbjct: 65  VVVKNLLSVEGLHWFLPNVIKNFSGFAPLGAILALVLGAGLAERVGLLPALMVKMASHVN 124

Query: 128 RKLTTFMVVFTGILSNTASELGYVVLIPLSAVIFHSLGRHPLAGLAAAFAGVSGGYSANL 187
            +  ++MV+F    S+ +S+   V++ P+ A+IF ++GRHP+AGL AA AGV  G++ANL
Sbjct: 125 ARYASYMVLFIAFFSHISSDAALVIMPPMGALIFLAVGRHPVAGLLAAIAGVGCGFTANL 184

Query: 188 FLGTIDPLLAGITQQAAQIIHPDYVVGPEANWFFMAASTFVIALIGYFVTEKIVEPQLGP 247
            + T D LL+GI+ +AA   +P   V    NW+FMA+S  V+ ++G  +T+KI+EP+LG
Sbjct: 185 LIVTTDVLLSGISTEAAAAFNPQMHVSVIDNWYFMASSVVVLTIVGGLITDKIIEPRLGQ 244

Query: 248 YQSDLSQEEKDIRHSNEITPLEYKGLIWAGVVFVALSALLAWSIVPADGILRHPETGLVA 307
           +Q +  ++ + +  S         GL  AGVV +   A +A  ++P +GILR P    V
Sbjct: 245 WQGNSDEKLQTLTESQRF------GLRIAGVVSLLFIAAIALMVIPQNGILRDPINHTVM 298

Query: 308 GSPFLKSIVVFIFLLFALPGIVYGRITRSLRGEREVVNAMAESMSTLGLYLXXXXXXXXX 367
            SPF+K IV  I L F +  + YG  TR++R + ++ + M E M  +  ++
Sbjct: 299 PSPFIKGIVPLIILFFFVVSLAYGIATRTIRRQADLPHLMIEPMKEMAGFIVMVFPLAQF 358

Query: 368 XXXXNWTNIGQYIAVKGAVFLKEVGLGGSVLFIGFILICAFINLMIGSASAQWAVTAPIF 427
               NW+N+G++IAV     L+  GL G   F+G  L+ +F+ + I S SA W++ APIF
```

```
Sbjct: 359 VAMFNWSNMGKFIAVGLTDILESSGLSGIPAFVGLALLSSFLCMFIASGSAIWSILAPIF 418

Query: 428 VPMLMLAGYAPEVIQAAYRIGDSVTNIITPMMSYFGLIMATVIKYKKDAGVGTLISMMLP 487
           VPM ML G+ P   Q  +RI DS     + P+  +  L +   + +YK DA +GT  S++LP
Sbjct: 419 VPMFMLLGFHPAFAQILFRIADSSVLPLAPVSPFVPLFLGFLQRYKPDAKLGTYYSLVLP 478

Query: 488 YSAFFLIAWIALFCIWVFVLGLPVGPG 514
           Y   FL+ W+ +   W +++GLP+GPG
Sbjct: 479 YPLIFLVVWLLMLLAW-YLVGLPIGPG 504
```

Based on this analysis, including the presence of several putative transmembrane domains and the predicted actinin-type actin-binding domain signature (shown in bold) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 17

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 141>:

```
  1  ..ACAGCCGGCG CAGCAGGTTn CnCGGTCTTC GTTTTCGTAA CGGACAGTCA
 51    GGTGGAGGTG TTCGGGAACA TCCAGACCGC AGTGGAAACA GGTTTTTTTC
101    ATGGCATTTC GGTTTCGTCT GTGTTTGGTG CGGCGGCACA AGACTCGGCA
151    ATgGCTTCGC GCAGTGCGTC TATACCGGTA TTTTCAGCAA CGGAAATGCG
201    GACGGcGgCA ATTTTTCCCG CAGCGTCGCG CCATATGCCC GTGTTTTgTT
251    CTTCAGACGG CAGCAGGTCG GTTTTGTTGT ACACCTTgAT GCACGGAaTA
301    TCGCCGGCAT GGATTTCTTG CAGTACGTTT TCCACGTCTT CAATCTGCTG
351    TCCGCTGTTC GGAGCGGCGG CATCGACGAC GTGCAGCAGC ACATCgGcTT
401    gCGCGGTTTC TTCCAGCGTG GCgGAAAAGG CGGAAATCAG TTTgTGCGGC
451    agATyGCTnA CGAATCCGAC GGTATCGGTC AGGATAATGC TGCATTCGGG
501    ACT..
```

This corresponds to the amino acid sequence <SEQ ID 142; ORF14>:

```
  1  ..TAGAAGXXVF VFVTDSQVEV FGNIQTAVET GFFHGISVSS VFGAAAQDSA
 51    MASRSASIPV FSATEMRTAA IFPAASRHMP VFCSSDGSRS VLLYTLMHGI
101    SPAWISCSTF STSSICCPLF GAAASTTCSS TSACAVSSSV AEKAEISLCG
151    RXLTNPTVSV RIMLHSG..
```

Computer analysis of this amino acid sequence gave the following results:

<u>Homology with a predicted ORF from *N.meningitidis* (strain A)</u>

ORF14 shows 94.0% identity over a 167aa overlap with an ORF (ORF14a) from strain A of *N. meningitidis:*

```
                                                     10        20        30
orf14.pep                                    TAGAAGXXVFVFVTDSQVEVFGNIQTAVET
                                             |:||||  |||||||:|::||||:| ||||
orf14a         GRQLGFLRVGGALFVITAQARVNNALCDCLTTGAAGFAVFVFVTDGQMQVFGNVQPAVET
               150       160       170       180       190       200

                       40        50        60        70        80        90
orf14.pep      GFFHGISVSSVFGAAAQDSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS
               |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
orf14a         GFFHGISVSSVFGAAAQYSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS
               210       220       230       240       250       260

                      100       110       120       130       140       150
orf14.pep      VLLYTLMHGISPAWISCSTFSTSSICCPLFGAAASTTCSSTSACAVSSSVAEKAEISLCG
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf14a         VLLYTLMHGISPAWISCSTFSTSSICCPLFGAAASTTCSSTSACAVSSSVAEKAEISLCG
               270       280       290       300       310       320
```

```
                     160
orf14.pep     RXLTNPTVSVRIMLHSG
              | ||||||||||||||| 
orf14a        RSLTNPTVSVRIMLHSGLMYSRRAVVSSVAKSWSFAYMPDLVSRLNRLDLPTLVX
             330       340       350       360       370       380
```

The complete length ORF14a nucleotide sequence <SEQ ID 143> is:

```
   1  ATGGAGGATT TGCAGGAAAT CGGGTTCGAT GTCGCCGCCG TAAAGGTAGG
  51  TCGGCAGCGC GAACATCATC GTCTGCATCA TCCCCAGCCC GGCAACGGCG
 101  AGGCGGACGA TGTATTGTTT GCGTTCTTTT TGGTTGGCGG CTTCGATTTT
 151  TTGCGCGTCA TAGGGTGCGG CGGTGTAGCC TATCTGCCTG ATTTTCAACA
 201  GAATGTCGGA AAGGCGGATT TTGCCGTCGT CCCAGACGAC GCGGCAGCGG
 251  TGCGTGCTGT AATTGAGGTC GATGCGGACG ATGCCGTCTG TACGCAAAAG
 301  CTGCTGTTCG ATCAGCCAGA CGCAGGCGGC GCAGGTGATG CCGCCGAGCA
 351  TTAAAACCGC CTCGCGCGTG CCGCCGTGGG TTTCCACAAA GTCGGACTGG
 401  ACTTCGGGCA GGTCGTACAG GCGGATTTGG TCGAGGATTT CTTGGGGCGG
 451  CAGCTCGGTT TTTTGCGCGT CGGCGGTGCG TTGTTTGTAA TAACTGCCCA
 501  AGCCCGCGTC AATAATGCTT TGTGCGACTG CCTGACAACC GGCGCAGCAG
 551  GTTTCGCGGT CTTCGTTTTC GTAACGGACG GTCAGATGCA GGTTTTCGGG
 601  AACGTCCAGC CCGCAGTGGA AACAGGTTTT TTTCATGGCA TTTCGGTTTC
 651  GTCTGTGTTT GGTGCGGCGG CACAATACTC GGCAATGGCT TCGCGCAGTG
 701  CGTCTATACC GGTATTTTCA GCAACGGAAA TGCGGACGGC CGCAATTTTT
 751  CCCGCAGCGT CGCGCCATAT GCCCGTGTTT TGTTCTTCAG ACGCCAGCAG
 801  GTCGGTTTTG TTGTACACCT TGATGCACGG AATATCGCCG GCATGCATTT
 851  CTTGCAGTAC GTTTTCCACG TCTTCAATCT GCTGTCCGCT GTTCGCAGCG
 901  GCGGCATCGA CGACGTGCAG CAGCACATCG GCTTGCGCGG TTTCTTCCAC
 951  CGTGGCGGAA AAGGCGGAAA TCAGTTTGTG CGGCAGATCG CTGACGAATC
1001  CGACGGTATC GGTCAGGATA ATGCTGCATT CGGGACTGAT GTACAGCCGC
1051  CGCGCCGTCG TGTCGAGTGT GGCGAAAAGC TGGTCTTTCG CATATATGCC
1101  CGACTTGGTC AGCCGGTTGA ACAGACTGGA TTTGCCGACA TTGGTATAG
```

This encodes a protein having amino acid sequence <SEQ ID 144>:

```
   1  MEDLQEIGFD VAAVKVGRQR EHHRLHHPQP GNGEADDVLF AFFLVGGFDF
  51  LRVIGCGGVA YLPDFQQNVG KADFAVVPDD AAAVRAVIEV DADDAVCTQK
 101  LLFDQPDAGG AGDAAEH*NR LARAAVGFHK VGLDFGQVVQ ADLVEDFLGR
 151  QLGFLRVGGA LFVITAQARV NNALCDCLTT GAAGFAVFVF VTDGQMQVFG
 201  NVQPAVETGF FHGISVSSVF GAAAQYSAMA SRSASIPVFS ATEMRTAAIF
 251  PAASRHMPVF CSSDGSRSVL LYTLMHGISP AWISCSTFST SSICCPLFGA
 301  AASTTCSSTS ACAVSSSVAE KAEISLCGRS LTNPTVSVRI MLHSGLMYSR
 351  RAVVSSVAKS WSFAYMPDLV SRLNRLDLPT LV*
```

It should be noted that this sequence includes a stop codon at position 118.

<u>Homology with a predicted ORF from *N.gonorrhoeae*</u>

ORF14 shows 89.8% identity over a 167aa overlap with a predicted ORF (ORF14.ng) from *N. gonorrhoeae:*

```
orf14.pep                                       TAGAAGXXVFVFVTDSQVEVFGNIQTAVET   30
                                                || |||  ||:||:|:|::||||:| ||||
orf14ng       GRQFGFFRVGGASFVITAQAGIDDALCDCLTADAAGFAVFAFVADGQMQVFGNVQPAVET  208

orf14.pep     GFFHGISVSSVFGAAAQDSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS   90
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
orf14ng       GFFHGISVSSVFGAAAQYSAMASRSASIPVFSATEMRTAAIFPAASRHMPVFCSSDGSRS  268

orf14.pep     VLLYTLMHGISPAWISCSTFSTSSICCPLFGAAASTTCSSTSACAVSSSVAEKAEISLCG  150
              |||||||||| ||||||||||||||||| |||||||||||||:|||:|||||||||||
orf14ng       VLLYTLMHGISWAWISCSTFSTSSICCPLFRAAASTTCSSTSACTVSSKVAEKAEISLCG  328

orf14.pep     RXLTNPTVSVRIMLHSG                                          167
              | |||||||||||||:|
orf14ng       RSLTNPTVSVRIMLHAGLMYSRRAVVSRVAKSWSFAYMPDLVSRLNRLDLPTLV  382
```

The complete length ORF14ng nucleotide sequence <SEQ ID 145> is predicted to encode a protein having amino acid sequence <SEQ ID 146>:

```
  1  MEDLQEIGFD VAAVKVGRQR EHHRLHHTQS GNGKADDVLF AFFLVGGFDF
 51  LRVIGCGGVA CLPDFQQNVG EADFAVVPDD AAAVRAVIEV DADDAVCAQK
101  LLFDQPDAGG AGNAAEHQHC FVRAIMGFHK VGLDFGQVVQ ADLVEDFLGR
151  QFGFFRVGGA SFVITAQAGI DDALCDCLTA DAAGFAVFAF VADGQMQVFG
201  NVQPAVETGF FHGISVSSVF GAAAQYSAMA SRSASIPVFS ATEMRTAAIF
251  PAASRHMPVF CSSDGSRSVL LYTLMHGISW AWISCSTFST SSICCPLFRA
301  AASTTCSSTS ACTVSSKVAE KAEISLCGRS LTNPTVSVRI MLHAGLMYSR
351  RAVVSRVAKS WSFAYMPDLV SRLNRLDLPT LV*
```

Based on the putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 18

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 147>:

```
  1  ..GGCCATTACT CCGACCGCAC TTGGAAGCCG CGTTTGGNCG GCCGCCGTCT
 51    GCCGTATCTG CTTTATGGCA CGCTGATTGC GGTTATTGTG ATGATTTTGA
101    TGCCGAACTC GGGCAGCTTC GGTTTCGGCT ATGCGTCGCT GGCGGCTTTG
151    TCGTTCGGCG CGCTGATGAT TGCGCTGTTA GACGTGTCGT CAAATATGGC
201    GATGCAGCCG TTTAAGATGA TGGTCGGCGA CATGGTCAAC GAGGAGCAGA
251    AAA.NTACGC CTACGGGATT CAAAGTTTCT TAGCAAATAC GGGCGCGGTC
301    GTGGCGGCGA TTCTGCCGTT TGTGTTTGCG TATATCGGTT TGGCGAACAC
351    CGCCGANAAA GGCGTTGTGC CGCAGACCGT GGTCGTGGCG TTTTATGTGG
401    GTGCGGCGTT GCTGGTGATT ACCAGCGCGT TCACGATTTT CAAAGTGAAG
451    GAATACGANC CGGAAACCTA CGCCCGTTAC CACGGCATCG ATGTCGCCGC
501    GAATCAGGAA AAAGCCAACT GGATCGCACT CTTAAAA.CC GCGC..
```

This corresponds to the amino acid sequence <SEQ ID 148; ORF16>:

```
  1  ..GHYSDRTWKP RLXGRRLPYL LYGTLIAVIV MILMPNSGSF GFGYASLAAL
 51    SFGALMIALL DVSSNMAMQP FKMMVGDMVN EEQKXYAYGI QSFLANTGAV
101    VAAILPFVFA YIGLANTAXK GVVPQTVVVA FYVGAALLVI TSAFTIFKVK
151    EYXPETYARY HGIDVAANQE KANWIALLKX A..
```

Further work revealed the complete nucleotide sequence <SEQ ID 149>:

```
   1  ATGTCGGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC
  51  AAAAAGCACG ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG
 101  CCTTTACCCT GCAAAGCTCG CAAATGAGCC GCATTTTTCA AACGCTAGGC
 151  GCAGACCCGC ACAATTTGGG CTGGTTTTTC ATCCTGCCGC CGCTGGCGGG
 201  GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCCGACCGC ACTTGGAAGC
 251  CGCGTTTGGG CGGCCGCCGT CTGCCGTATC TGCTTTATGG CACGCTGATT
 301  GCGGTTATTG TGATGATTTT GATGCCGAAC TCGGGCAGCT TCGGTTTCGG
 351  CTATGCGTCG CTGGCGGCTT TGTCGTTCGG CGCGCTGATG ATTGCGCTGT
 401  TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTTAAGAT GATGGTCGGC
 451  GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT
 501  CTTAGCAAAT ACGGGCGCGG TCGTGGCGGC GATTCTGCCG TTTGTGTTTG
 551  CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCGCAGACC
 601  GTGGTCGTGG CGTTTTATGT GGGTGCGGCG TTGCTGGTGA TTACCAGCGC
 651  GTTCACGATT TTCAAAGTGA AGGAATACGA TCCGGAAACC TACGCCCGTT
 701  ACCACGGCAT CGATGTCGCC GCGAATCAGG AAAAAGCCAA CTGGATCGAA
 751  CTCTTGAAAA CCGCGCCTAA GGCGTTTTGG ACGGTTACTT TGGTGCAATT
 801  CTTCTGCTGG TTCGCCTTCC AATATATGTG GACTTACTCG GCAGGCGCGA
 851  TTGCGGAAAA CGTCTGGCAC ACCACCGATG CGTCTTCCGT AGGTTATCAG
 901  GAGGCGGGTA ACTGGTACGG CGTTTTGGCG GCGGTGCAGT CGGTTGCGGC
 951  GGTGATTTGT TCGTTTGTAT TGGCGAAAGT GCCGAATAAA TACCATAAGG
1001  CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCCGTT
1051  TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGTCTTATA CCTTAATCGG
1101  CATCGCTTGG GCGGGCATTA TCACTTATCC GCTGACGATT GTGACCAACG
1151  CCTTGTCGGG CAAGCATATG GGCACTTACT TGGGCTTGTT TAACGGCTCT
1201  ATCTGTATGC CTCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTTCC
1251  TATGCTGGGC GGCTTGCAGG CCACTATGTT CTTGGTAGGG GGCGTCGTCC
1301  TGCTGCTGGG CGCGTTTTCC GTGTTCCTGA TTAAAGAAAC ACACGGCGGG
1351  GTTTGA
```

This corresponds to the amino acid sequence <SEQ ID 150; ORF16-1>:

```
  1  MSEYTPQTAK QGLPALAKST IWMLSFGFLG VQTAFTLQSS QMSRIFQTLG
 51  ADPHNLGWFF ILPPLAGMLV QPIVGHYSDR TWKPRLGGRR LPYLLYGTLI
101  AVIVMILMPN SGSFGFGYAS LAALSFGALM IALLDVSSNM AMQPFKMMVG
151  DMVNEEQKGY AYGIQSFLAN TGAVVAAILP FVFAYIGLAN TAEKGVVPQT
201  VVVAFYVGAA LLVITSAFTI FKVKEYDPET YARYHGIDVA ANQEKANWIE
251  LLKTAPKAFW TVTLVQFFCW FAFQYMWTYS AGAIAENVWH TTDASSVGYQ
301  EAGNWYGVLA AVQSVAAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFFSV
351  FFIGNQYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGLFNGS
401  ICMPQIVASL LSFVLFPMLG GLQATMFLVG GVVLLLGAFS VFLIKETHGG
451  V*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF16 shows 96.7% identity over a 181aa overlap with an ORF (ORF16a) from strain A of *N. meningitidis:*

```
                                                10        20        30
orf16.pep                             GHYSDRTWKPRLXGRRLPYLLYGTLIAVIV
                                      |||||||||||| ||||||||||||||||
orf16a     IFQTLGADPHSLGWFFILPPLAGMLVQPIVGHYSDRTWKPRLGGRRLPYLLYGTLIAVIV
                50        60        70        80        90       100

                   40        50        60        70        80        90
orf16.pep  MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKXYAYGI
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
orf16a     MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKGYAYGI
                110       120       130       140       150       160

                  100       110       120       130       140       150
orf16.pep  QSFLANTGAVVAAILPFVFAYIGLANTAXKGVVPQTVVVAFYVGAALLVITSAFTIFKVK
           |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
orf16a     QSFLANTGAVVAAILPFVFAYIGLANTAEKGVVPQTVVVAFYVGAALLVITSAFTIFKVK
                170       180       190       200       210       220

                  160       170       180
orf16.pep  EYXPETYARYHGIDVAANQEKANWIALLKXA
           || ||||||||||||||||||||||| |||:|
orf16a     EYNPETYARYHGIDVAANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWTYSAGAI
                230       240       250       260       270       280

orf16a     AENVWHTTDASSVGYQEAGNWYGVLAAVQSVAAVICSFVLAKVPNKYHKAGYFGCLALGA
                290       300       310       320       330       340
```

The complete length ORF16a nucleotide sequence <SEQ ID 151> is:

```
  1  ATGTCGGAAT ATACGCCTCA AACAGCAAAA CAAGGTTTGC CCGCGCTGGC
 51  AAAAAGCACG ATTTGGATGC TCAGTTTCGG CTTTCTCGGC GTTCAGACGG
101  CCTTTACCCT GCAAAGCTCG CAGATGAGCC GCATCTTCCA GACGCTCGGT
151  GCCGATCCGC ACAGCCTCGG CTGGTTCTTT ATCCTGCCGC CGCTGGCGGG
201  GATGCTGGTG CAGCCGATTG TCGGCCATTA CTCCGACCGC ACTTGGAAGC
251  CGCGTTTGGG CGGCCGCCGT CTGCCGTATC TGCTTTATGG CACGCTGATT
301  GCGGTTATTG TGATGATTTT GATGCCGAAC TCGGGCAGCT TCGGTTTCGG
351  CTATGCGTCG CTGGCGGCTT TGTCGTTCGG CGCGCTGATG ATTGCGCTGT
401  TAGACGTGTC GTCAAATATG GCGATGCAGC CGTTTAAGAT GATGGTCGGC
451  GACATGGTCA ACGAGGAGCA GAAAGGCTAC GCCTACGGGA TTCAAAGTTT
501  CTTAGCGAAT ACGGGCGCGG TCGTGGCGGC GATTCTGCCG TTTGTGTTTG
551  CGTATATCGG TTTGGCGAAC ACCGCCGAGA AAGGCGTTGT GCCGCAGACC
601  GTGGTCGTGG CGTTTTATGT GGGTGCGGCG TTGCTGGTGA TTACCAGCGC
651  GTTCACGATT TTCAAAGTGA AGGAATACAA TCCGGAAACC TACGCCCGTT
701  ACCACGGCAT CGATGTCGCC GCGAATCAGG AAAAAGCCAA CTGGATCGAA
751  CTCTTGAAAA CCGCGCCTAA GGCGTTTTGG ACGGTTACTT TGGTGCAATT
801  CTTCTGCTGG TTCGCCTTCC AATATATGTG GACTTACTCG GCAGGCGCGA
851  TTGCGGAAAA CGTCTGGCAC ACCACCGATG CGTCTTCCGT AGGTTATCAG
901  GAGGCGGGTA ACTGGTACGG CGTTTTGGCG GCGGTGCAGT CGGTTGCGGC
951  GGTGATTTGT TCGTTTGTAT TGGCGAAAGT GCCGAATAAA TACCATAAGG
```

```
1001  CGGGTTATTT CGGCTGTTTG GCTTTGGGCG CGCTCGGCTT TTTCTCCGTT
1051  TTCTTCATCG GCAACCAATA CGCGCTGGTG TTGTCTTATA CCTTAATCGG
1101  CATCGCTTGG GCGGGCATTA TCACTTATCC GCTGACGATT GTGACCAACG
1151  CCTTGTCGGG CAAGCATATG GGCACTTACT TGGGCCTGTT TAACGGCTCT
1201  ATCTGTATGC CGCAAATCGT CGCTTCGCTG TTGAGTTTCG TGCTTTTCCC
1251  TATGCTGGGC GGCTTGCAGG CCACTATGTT CTTGGTAGGG GGCGTCGTCC
1301  TGCTGCTGGG CGCGTTTTCC GTGTTCCTGA TTAAAGAAAC ACACGGCGGG
1351  GTTTGA
```

This encodes a protein having amino acid sequence <SEQ ID 152>:

```
  1  MSEYTPQTAK QGLPALAKST IWMLSFGFLG VQTAFTLQSS QMSRIFQTLG
 51  ADPHSLGWFF ILPPLAGMLV QPIVGHYSDR TWKPRLGGRR LPYLLYGTLI
101  AVIVMILMPN SGSFGFGYAS LAALSFGALM IALLDVSSNM AMQPFKMMVG
151  DMVNEEQKGY AYGIQSFLAN TGAVVAAILP FVFAYIGLAN TAEKGVVPQT
201  VVVAFYVGAA LLVITSAFTI FKVKEYNPET YARYHGIDVA ANQEKANWIE
251  LLKTAPKAFW TVTLVQFFCW FAFQYMWTYS AGAIAENVWH TTDASSVGYQ
301  EAGNWYGVLA AVQSVAAVIC SFVLAKVPNK YHKAGYFGCL ALGALGFFSV
351  FFIGNQYALV LSYTLIGIAW AGIITYPLTI VTNALSGKHM GTYLGLFNGS
401  ICMPQIVASL LSFVLFPMLG GLQATMFLVG GVVLLLGAFS VFLIKETHGG
451  V*
```

ORF16a and ORF16-1 show 99.6% identity in 451 aa overlap:

```
                      10        20        30        40        50        60
orf16a.pep   MSEYTPQTAKQGLPALAKSTIWMLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHSLGWFF
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
orf16-1      MSEYTPQTAKQGLPALAKSTIWMLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHNLGWFF
                      10        20        30        40        50        60

                      70        80        90       100       110       120
orf16a.pep   ILPPLAGMLVQPIVGHYSDRTWKPRLGGRRLPYLLYGTLIAVIVMILMPNSGSFGFGYAS
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf16-1      ILPPLAGMLVQPIVGHYSDRTWKPRLGGRRLPYLLYGTLIAVIVMILMPNSGSFGFGYAS
                      70        80        90       100       110       120

                     130       140       150       160       170       180
orf16a.pep   LAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKGYAYGIQSFLANTGAVVAAILP
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf16-1      LAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKGYAYGIQSFLANTGAVVAAILP
                     130       140       150       160       170       180

                     190       200       210       220       230       240
orf16a.pep   FVFAYIGLANTAEKGVVPQTVVVAFYVGAALLVITSAFTIFKVKEYNPETYARYHGIDVA
             |||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
orf16-1      FVFAYIGLANTAEKGVVPQTVVVAFYVGAALLVITSAFTIFKVKEYDPETYARYHGIDVA
                     190       200       210       220       230       240

                     250       260       270       280       290       300
orf16a.pep   ANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWTYSAGAIAENVWHTTDASSVGYQ
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf16-1      ANQEKANWIELLKTAPKAFWTVTLVQFFCWFAFQYMWTYSAGAIAENVWHTTDASSVGYQ
                     250       260       270       280       290       300

                     310       320       330       340       350       360
orf16a.pep   EAGNWYGVLAAVQSVAAVICSFVLAKVPNKYHKAGYFGCLALGALGFFSVFFIGNQYALV
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf16-1      EAGNWYGVLAAVQSVAAVICSFVLAKVPNKYHKAGYFGCLALGALGFFSVFFIGNQYALV
                     310       320       330       340       350       360

                     370       380       390       400       410       420
orf16a.pep   LSYTLIGIAWAGIITYPLTIVTNALSGKHMGTYLGLFNGSICMPQIVASLLSFVLFPMLG
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf16-1      LSYTLIGIAWAGIITYPLTIVTNALSGKHMGTYLGLFNGSICMPQIVASLLSFVLFPMLG
                     370       380       390       400       410       420

                     430       440       450
orf16a.pep   GLQATMFLVGGVVLLLGAFSVFLIKETHGGVX
             |||||||||||||||||||||||||||||||
orf16-1      GLQATMFLVGGVVLLLGAFSVFLIKETHGGVX
                     430       440       450
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF16 shows 93.9% identity over a 181aa overlap with a predicted ORF (ORF16.ng) from *N. gonorrhoeae:*

```
orf16.pep                                  GHYSDRTWKPRLXGRRLPYLLYGTLIAVIV   30
                                           |:|||||||||| |||||||||||||||||
orf16ng     HFSNARRRPAQFGLVFHPAAAGGDAGSADSGYYSDRTWKPRLGGRRLPYLLYGTLIAVIV  131

orf16.pep   MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKXYAYGI   90
            |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
orf16ng     MILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMAMQPFKMMVGDMVNEEQKSYAYGI  191

orf16.pep   QSFLANTGAVVAAILPFVFAYIGLANTAXKGVVPQTVVVAFYVGAALLVITSAFTIFKVK  150
            ||||||| |||||||||||||||||||| |||||||||||||||||||:||||||| |||
orf16ng     QSFLANTDAVVAAILPFVFAYIGLANTAEKGVVPQTVVVAFYVGAALLIITSAFTISKVK  251

orf16.pep   EYXPETYARYHGIDVAANQEKANWIALLKXA                               181
            || |||||||||||||||||||||: |||:|
orf16ng     EYDPETYARYHGIDVAANQEKANWFELLKTAPKVFWTVTPVQFFCWFAFRYMWTYSAGAI  311
```

The complete length ORF16ng nucleotide sequence <SEQ ID 153> is:

```
   1  ATGATAGGGG ATCGCCGCGC CGGCAACCAT TTCGGATTTT CCAAAGCAAA
  51  TACTTTTCAA ATCAAAAAAA AGGATTTACT TTATGTCGGA ATATACGCCT
 101  CAAACAGCAA AACAAGGTTT GCCCGCGCCG GCAAAAAGCA CGATTTGGAT
 151  GTTGAGCTTC GGCTATCTCG GCGTTCAGAC GGCCTTTACC CTGCAAAGCT
 201  CGCAGATGAG CCGCATTTTT CAAACGCTAG GCGCAGACCC GCACAATTTG
 251  GGCTGGTTTT TCATCCTGCC GCCGCTGGCG GGGATGCTGG TTCAGCCGAT
 301  AGTGGCTACT ACTCAGACCG CACTTGGAAG CCGCGCTTGG GCGGCCGCCG
 351  CCTGCCGTAT CTGCTTTACG GCACGCTGAT TGCGGTCATC GTGATGATTT
 401  TGATGCCGAA CTCGGGCAGC TTCGGTTTCG GCTATGCGTC GCTGGCGGCC
 451  TTGTCGTTCG GCGCGCTGAT GATTGCGCTG TTGGACGTGT CGTCGAATAT
 501  GGCGATGCAG CCGTTTAAGA TGATGGTCGG CGATATGGTC AACGAGGAGC
 551  AGAAAAGCTA CGCCTACGGG ATTCAAAGTT TCTTAGCGAA TACGGACGCG
 601  GTTGTGGCAG CGATTCTGCC GTTTGTGTTC GCGTATATCG GTTTGGCGAA
 651  CACTGCCGAG AAAGGCGTTG TGCCACAAAC CGTGGTCGTA GCATTCTATG
 701  TGGGTGCGGC GTTACTGATT ATTACCAGTG CGTTCACAAT CTCCAAAGTC
 751  AAAGAATACG ACCCGGAAAC CTACGCCCGT TACCACGGCA TCGATGTCGC
 801  CGCGAATCAG GAAAAAGCCA ACTGGTTCGA ACTCTTAAAA ACCGCGCCTA
 851  AAGTGTTTTG GACGGTTACT CCGGTACAGT TTTTCTGCTG GTTCGCCTTC
 901  CGGTATATGT GGACTTACTC GGCAGGCGCG ATTGCAGAAA ACGTCTGGCA
 951  CACTACCGAT GCGTCTTCCG TAGGCCATCA GGAGGCGGGC AACCGGTACG
1001  GCGTTTTGGC GGCGGTGTAG
```

This encodes a protein having amino acid sequence <SEQ ID 154>:

```
   1  MIGDRRAGNH FGFSKANTFQ IKKKDLLYVG IYASNSKTRF ARAGKKHDLD
  51  VELRLSRRSD GLYPAKLADE PHFSNARRRP AQFGLVFHPA AAGGDAGSAD
 101  SGYYSDRTWK PRLGGRRLPY LLYGTLIAVI VMILMPNSGS FGFGYASLAA
 151  LSFGALMIAL LDVSSNMAMQ PFKMMVGDMV NEEQKSYAYG IQSFLANTDA
 201  VVAAILPFVF AYIGLANTAE KGVVPQTVVV AFYVGAALLI ITSAFTISKV
 251  KEYDPETYAR YHGIDVAANQ EKANWFELLK TAPKVFWTVT PVQFFCWFAF
 301  RYMWTYSAGA IAENVWHTTD ASSVGHQEAG NRYGVLAAV*
```

ORF16ng and ORF16-1 show 89.3% identity in 261 aa overlap:

```
                      30        40        50        60        70        80
orf16-1.pep   MLSFGFLGVQTAFTLQSSQMSRIFQTLGADPHNLGWFFILPPLAGMLVQPI-VGHYSDRT
                                            | ::|  |  |  ||  :     |:|||||
orf16ng       DVELRLSRRSDGLYPAKLADEPHFSNARRRPAQFGLVF-HPAAAGGDAGSADSGYYSDRT
              50        60        70        80          90        100

                      90       100       110       120       130       140
orf16-1.pep   WKPRLGGRRLPYLLYGTLIAVIVMILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMA
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf16ng       WKPRLGGRRLPYLLYGTLIAVIVMILMPNSGSFGFGYASLAALSFGALMIALLDVSSNMA
              110       120       130       140       150       160
```

```
                         150       160       170       180       190       200
orf16-1.pep   MQPFKMMVGDMVNEEQKGYAYGIQSFLANTGAVVAAILPFVFAYIGLANTAEKGVVPQTV
              |||||||||||||||||:|||||||||||| |||||||||||||||||||||||||||||
orf16ng       MQPFKMMVGDMVNEEQKSYAYGIQSFLANTDAVVAAILPFVFAYIGLANTAEKGVVPQTV
            170       180       190       200       210       220

                         210       220       230       240       250       260
orf16-1.pep   VVAFYVGAALLVITSAFTIFKVKEYDPETYARYHGIDVAANQEKANWIELLKTAPKAFWT
              |||||||||||:||||||| |||||||||||||||||||||||||||:||||||||:|||
orf16ng       VVAFYVGAALLIITSAFTISKVKEYDPETYARYHGIDVAANQEKANWFELLKTAPKVFWT
            230       240       250       260       270       280

                         270       280       290       300       310       320
orf16-1.pep   VTLVQFFCWFAFQYMWTYSAGAIAENVWHTTDASSVGYQEAGNWYGVLAAVQSVAAVICS
              || |||||||||:||||||||||||||||||||||||:||||| |||||||
orf16ng       VTPVQFFCWFAFRYMWTYSAGAIAENVWHTTDASSVGHQEAGNRYGVLAAVX
            290       300       310       320       330       340
```

Based on this analysis, including the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 19

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 155>:

```
  1  ATGTTGTTCC GTAAAACGAC CGCCGCCGTT TTGGCGCATA CCTTGATGCT
 51  GAACGGCTGT ACGTTGATGT TGTGGGGAAT GAACAACCCG GTCAGCGAAA
101  CAATCACCCG NAAACACGTT GNCAAAGACC AAATCCGNGN CTTCGGTGTG
151  GTTGCCGAAG ACAATGCCCA ATTGGAAAAG GGCAGCCTGG TGATGATGGG
201  CGGAAAATAC TGGTTCGTCG TCAATCCCGA AGATTCGGCG AA.NTGACGG
251  GNATTTTGAN GGCAGGGCTG GACAAACCCT TCCAAATAGT TNAGGATACC
301  CCGAGCTATG C.TGCCACCA AGCCCTGCCG GTCAAACTCG GATCGNCTGG
351  CAGCCAGAAT...
```

This corresponds to the amino acid sequence <SEQ ID 156; ORF28>:

```
  1  MLFRKTTAAV LAHTLMLNGC TLMLWGMNNP VSETITRKHV XKDQIRXFGV
 51  VAEDNAQLEK GSLVMMGGKY WFVVNPEDSA XXTGILXAGL DKPFQIVXDT
101  PSYXCHQALP VKLGSXGSQN...
```

Further work revealed the complete nucleotide sequence <SEQ ID 157>:

```
  1  ATGTTGTTCC GTAAAACGAC CGCCGCCGTT TTGGCGGCAA CCTTGATGCT
 51  GAACGGCTGT ACGTTGATGT TGTGGGGAAT GAACAACCCG GTCAGCGAAA
101  CAATCACCCG CAAACACGTT GACAAAGACC AAATCCGCGC CTTCGGTGTG
151  GTTGCCGAAG ACAATGCCCA ATTGGAAAAG GGCAGCCTGG TGATGATGGG
201  CGGAAAATAC TGGTTCGTCG TCAATCCCGA AGATTCGGCG AAGCTGACGG
251  GCATTTTGAA GGCAGGGCTG GACAAACCCT TCCAAATAGT TGAGGATACC
301  CCGAGCTATG CTCGCCACCA AGCCCTGCCG GTCAAACTCG AATCGCCTGG
351  CAGCCAGAAT TTCAGTACCG AAGGCCTTTG CCTGCGCTAC GATACCGACA
401  AGCCTGCCGA CATCGCCAAG CTGAAACAGC TCGGGTTTGA AGCGGTCAAA
451  CTCGACAATC GGACCATTTA CACGCGCTGC GTATCCGCCA AAGGCAAATA
501  CTACGCCACA CCGCAAAAAC TGAACGCCGA TTACCATTTT GAGCAAAGTG
551  TGCCTGCCGA TATTTATTAC ACGGTTACTG AAGAACATAC CGACAAATCC
601  AAGCTGTTTG CAAATATCTT ATATACGCCC CCCTTTTTGA TACTGGATGC
651  GGCGGGCGCG GTACTGGCCT TGCCTGCGGC GGCTCTGGGT GCGGTCGTGG
701  ATGCCGCCCG CAAATGA
```

This corresponds to the amino acid sequence <SEQ ID 158; ORF28-1>:

```
  1  MLFRKTTAAV LAATLMLNGC TLMLWGMNNP VSETITRKHV DKDQIRAFGV
 51  VAEDNAQLEK GSLVMMGGKY WFVVNPEDSA KLTGILKAGL DKPFQIVEDT
101  PSYARHQALP VKLESPGSQN FSTEGLCLRY DTDKPADIAK LKQLGFEAVK
151  LDNRTIYTRC VSAKGKYYAT PQKLNADYHF EQSVPADIYY TVTEEHTDKS
```

```
201  KLFANILYTP PFLILDAAGA VLALPAAALG AVVDAARK*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF28 shows 79.2% identity over a 120aa overlap with an ORF (ORF28a) from strain A of *N. meningitidis:*

```
                       10        20        30        40        50        60
orf28.pep      MLFRKTTAAVLAHTLMLNGCTLMLWGMNNPVSETITRKHVXKDQIRXFGVVAEDNAQLEK
               ||||||||||| ||||||||:|:||||:| |||  :|||| ||||| ||||||||||||
orf28a         MLFRKTTAAVLAATLMLNGCTVMMWGMNSPFSETTARKHVDKDQIRAFGVVAEDNAQLEK
                       10        20        30        40        50        60

                       70        80        90       100       110       120
orf28.pep      GSLVMMGGKYWFVVNPEDSAXXTGILXAGLDKPFQIVXDTPSYXCHQALPVKLGSXGSQN
               ||||||||||||||||||||   |||| ||||| ||:|  :| :  :||||||| | :|||
orf28a         GSLVMMGGKYWFVVNPEDSAKLTGILKAGLDKQFQMVEPNPRFA-YQALPVKLESPASQN
                       70        80        90       100       110

orf28a         FSTEGLCLRYDTDRPADIAKLKQLEFEAVELDNRTIYTRCVSAKGKYYATPQKLNADYHF
            120       130       140       150       160       170
```

The complete length ORF28a nucleotide sequence <SEQ ID 159> is:

```
  1  ATGTTGTTCC GTAAAACGAC CGCCGCCGTT TTGGCGGCAA CCTTGATGTT
 51  GAACGGCTGT ACGGTAATGA TGTGGGGTAT GAACAGCCCG TTCAGCGAAA
101  CGACCGCCCG CAAACACGTT GACAAGGACC AAATCCGCGC CTTCGGTGTG
151  GTTGCCGAAG ACAATGCCCA ATTGGAAAAG GGCAGCCTGG TGATGATGGG
201  CGGGAAATAC TGGTTCGTCG TCAATCCTGA AGATTCGGCG AAGCTGACGG
251  GCATTTTGAA GGCCGGGTTG GACAAGCAGT TTCAAATGGT TGAGCCCAAC
301  CCGCGCTTTG CCTACCAAGC CCTGCCGGTC AAACTCGAAT CGCCCGCCAG
351  CCAGAATTTC AGTACCGAAG GCCTTTGCCT GCGCTACGAT ACCGACAGAC
401  CTGCCGACAT CGCCAAGCTG AAACAGCTTG AGTTTGAAGC GGTCGAACTC
451  GACAATCGGA CCATTTACAC GCGCTGCGTC TCCGCCAAAG GCAAATACTA
501  CGCCACACCG CAAAAACTGA ACGCCGATTA TCATTTTGAG CAAAGTGTGC
551  CTGCCGATAT TTATTACACG GTTACGAAAA AACATACCGA CAAATCCAAG
601  TTGTTTGAAA ATATTGCATA TACGCCCACC ACGTTGATAC TGGATGCGGT
651  GGGCGCGGTG CTGGCCTTGC CTGTCGCGGC GTTGATTGCA GCCACGAATT
701  CCTCAGACAA ATGA
```

This encodes a protein having amino acid sequence <SEQ ID 160>:

```
  1  MLFRKTTAAV LAATLMLNGC TVMMWGMNSP FSETTARKHV DKDQIRAFGV
 51  VAEDNAQLEK GSLVMMGGKY WFVVNPEDSA KLTGILKAGL DKQFQMVEPN
101  PRFAYQALPV KLESPASQNF STEGLCLRYD TDRPADIAKL KQLEFEAVEL
151  DNRTIYTRCV SAKGKYYATP QKLNADYHFE QSVPADIYYT VTKKHTDKSK
201  LFENIAYTPT TLILDAVGAV LALPVAALIA ATNSSDK*
```

ORF28a and ORF28-1 show 86.1% identity in 238 aa overlap:

```
                       10        20        30        40        50        60
orf28a.pep     MLFRKTTAAVLAATLMLNGCTVMMWGMNSPFSETTARKHVDKDQIRAFGVVAEDNAQLEK
               ||||||||||||||||||||:|:||||:| |||  :|||||||||||||||||||||||
orf28-1        MLFRKTTAAVLAATLMLNGCTLMLWGMNNPVSETITRKHVDKDQIRAFGVVAEDNAQLEK
                       10        20        30        40        50        60

                       70        80        90       100       110      119
orf28a.pep     GSLVMMGGKYWFVVNPEDSAKLTGILKAGLDKQFQMVEPNPRFA-YQALPVKLESPASQN
               ||||||||||||||||||||||||||||||||| ||:|| :| :| :||||||||||:|||
orf28-1        GSLVMMGGKYWFVVNPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPGSQN
                       70        80        90       100       110       120

            120       130       140       150       160       170      179
orf28a.pep     FSTEGLCLRYDTDRPADIAKLKQLEFEAVELDNRTIYTRCVSAKGKYYATPQKLNADYHF
               ||||||||||||:|||||||||| ||||:|||||||||||||||||||||||||||||||
orf28-1        FSTEGLCLRYDTDKPADIAKLKQLGFEAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF
                       130       140       150       160       170       180
```

```
            180       190       200       210       220       230
orf28a.pep  EQSVPADIYYTVTKKHTDKSKLFENIAYTPTTLILDAVGAVLALPVAALIAATNSSDKX
            |||||||||||||::|||||||| || |||  |||||:|||||||:||| |::::: ||
orf28-1     EQSVPADIYYTVTEEHTDKSKLFANILYTPPFLILDAAGAVLALPAAALGAVVDAARKX
                     190       200       210       220       230
```

<u>Homology with a predicted ORF from *N.gonorrhoeae*</u>

ORF28 shows 84.2% identity over a 120aa overlap with a predicted ORF (ORF28.ng) from *N. gonorrhoeae:*

```
orf28.pep   MLFRKTTAAVLAHTLMLNGCTLMLWGMNNPVSETITRKHVXKDQIRXFGVVAEDNAQLEK   60
            |||||||||||| ||:|||||:|| |||||||:||||||| ||||| |||||||||||||
orf28ng     MLFRKTTAAVLAATLILNGCTMMLRGMNNPVSQTITRKHVDKDQIRAFGVVAEDNAQLEK   60

orf28.pep   GSLVMMGGKYWFVVNPEDSAXXTGILXAGLDKPFQIVXDTPSYXCHQALPVKLGSXGSQN  120
            ||||||||||||:|||||||  ||:| |||||||||| |||||  |||||||: : ||||
orf28ng     GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSQN  120
```

The complete length ORF28ng nucleotide sequence <SEQ ID 161> is

```
  1  ATGTTGTTCC GTAAAACGAC CGCCGCCGTT TTGGCGGCAA CCTTGATACT
 51  GAACGGCTGT ACGATGATGT TGCGGGGGAT GAACAACCCG GTCAGCCAAA
101  CAATCACCCG CAAACACGTT GACAAAGACC AAATCCGCGC CTTCGGTGTG
151  GTTGCCGAAG ACAATGCCCA ATTGGAAAAG GGCAGCCTGG TGATGATGGG
201  CGGGAAATAC TGGTTCGCCG TCAATCCCGA AGATTCGGCG AAGCTGACGG
251  GCCTTTTGAA GGCCGGGTTG GACAAGCCCT TCCAAATAGT TGAGGATACC
301  CCGAGCTATG CCCGCCACCA AGCCCTGCCG GTCAAATTCG AAGCGCCCGG
351  CAGCCAGAAT TTCAGTACCG GAGGTCTTTG CCTGCGCTAT GATACCGGCA
401  GACCTGACGA CATCGCCAAG CTGAAACAGC TTGAGTTTAA AGCGGTCAAA
451  CTCGACAATC GGACCATTTA CACGCGCTGC GTATCCGCCA AAGGCAAATA
501  CTACGCCACG CCGCAAAAAC TGAACGCCGA TTATCATTTT GAGCAAAGTG
551  TGCCCGCCGA TATTTATTAT ACGGTTACTG AAAAACATAC CGACAAATCC
601  AAGCTGTTTG GAAATATCTT ATATACGCCC CCCTTGTTGA TATTGGATGC
651  GGCGGCCGCG GTGCTGGTCT TGCCTATGGC TCTGATTGCA GCCGCGAATT
701  CCTCAGACAA ATGA
```

This encodes a protein having amino acid sequence <SEQ ID 162>:

```
  1  MLFRKTTAAV LAATLILNGC TMMLRGMNNP VSQTITRKHV DKDQIRAFGV
 51  VAEDNAQLEK GSLVMMGGKY WFAVNPEDSA KLTGLLKAGL DKPFQIVEDT
101  PSYARHQALP VKFEAPGSQN FSTGGLCLRY DTGRPDDIAK LKQLEFKAVK
151  LDNRTIYTRC VSAKGKYYAT PQKLNADYHF EQSVPADIYY TVTEKHTDKS
201  KLFGNILYTP PLLILDAAAA VLVLPMALIA AANSSDK*
```

ORF28ng and ORF28-1 share 90.0% identity in 231 aa overlap:

```
                     10        20        30        40        50        60
orf28-1.pep  MLFRKTTAAVLAATLMLNGCTLMLWGMNNPVSETITRKHVDKDQIRAFGVVAEDNAQLEK
             |||||||||||||||:|||||:|| |||||||:|||||||||||||||||||||||||||
orf28ng      MLFRKTTAAVLAATLILNGCTMMLRGMNNPVSQTITRKHVDKDQIRAFGVVAEDNAQLEK
                     10        20        30        40        50        60

                     70        80        90       100       110       120
orf28-1.pep  GSLVMMGGKYWFVVNPEDSAKLTGILKAGLDKPFQIVEDTPSYARHQALPVKLESPGSQN
             ||||||||||||:|||||||||||:|||||||||||||||||||||||||||:|:|||||
orf28ng      GSLVMMGGKYWFAVNPEDSAKLTGLLKAGLDKPFQIVEDTPSYARHQALPVKFEAPGSQN
                     70        80        90       100       110       120

                    130       140       150       160       170       180
orf28-1.pep  FSTEGLCLRYDTDKPADIAKLKQLGFEAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF
             ||| |||||||| :| |||||||| |:|||||||||||||||||||||||||||||||||
orf28ng      FSTGGLCLRYDTGRPDDIAKLKQLEFKAVKLDNRTIYTRCVSAKGKYYATPQKLNADYHF
                    130       140       150       160       170       180

                    190       200       210       220       230      239
orf28-1.pep  EQSVPADIYYTVTEEHTDKSKLFANILYTPPFLILDAAGAVLALPAAALGAVVDAARKX
             ||||||||||||||:||||||||:|||||||:||||||:|||:|| |  ::|:
orf28ng      EQSVPADIYYTVTEKHTDKSKLFGNILYTPPLLILDAAAAVLVLPMALIAAANSSDKX
```

```
              190       200       210       220       230
```

Based on this analysis, including the presence of a putative transmembrane domain in the gonococcal protein, it was predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF28-1 (24kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 6A shows the results of affinity purification of the GST-fusion protein, and Figure 6B shows the results of expression of the His-fusion in *E.coli*. Purified GST-fusion protein was used to immunise mice, whose sera were used for ELISA, which gave a positive result. These experiments confirm that ORF28-1 is a surface-exposed protein, and that it may be a useful immunogen.

Example 20

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 163>:

```
  1  ..GTCAGTCCTG TACTGCCTAT TACACACGAA CGGACAGGGT TTGAAGGTGT
 51    TATCGGTTAT GAAACCCATT TTTCAGGGCA CGGACATGAA GTACACAGTC
101    CGTTCGATCA TCATGATTCA AAAAGCACTT CTGATTTCAG CGGCGGTGTA
151    GACGGCGGTT TTACTGTTTA CCAACTTCAT CGAACATGGT CGGAAATCCA
201    TCCGGAGGAT GAATATGACG GGCCGCAAGC AGCG.ATTAT CCGCCCCCCG
251    GAGGAGCAAG GGATATATAC AGCTATTATG TCAAAGGAAC TTCAACAAAA
301    ACAAAGACTA GTATTGTCCC TCAAGCCCCA TTTTCAGACC GTTGGCTAGA
351    AGAAAATGCC GGTGCCGCCT CTGGT..
```

This corresponds to the amino acid sequence <SEQ ID 164; ORF29>:

```
  1  ..VSPVLPITHE RTGFEGVIGY ETHFSGHGHE VHSPFDHHDS KSTSDFSGGV
 51    DGGFTVYQLH RTWSEIHPED EYDGPQAAXY PPPGGARDIY SYYVKGTSTK
101    TKTSIVPQAP FSDRWLEENA GAASG..
```

Further work revealed the complete nucleotide sequence <SEQ ID 165>:

```
   1  ATGAATTTGC CTATTCAAAA ATTCATGATG CTGTTTGCAG CAGCAATATC
  51  GTTGCTGCAA ATCCCCATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC
 101  GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGTAA ATACCATCTG
 151  TTTGGTAATG CTCGCGGCAG TGTTAAAAAG CGGGTTTACG CCGTCCAGAC
 201  ATTTGATGCA ACTGCGGTCA GTCCTGTACT GCCTATTACA CACGAACGGA
 251  CAGGGTTTGA AGGTGTTATC GGTTATGAAA CCCATTTTTC AGGGCACGGA
 301  CATGAAGTAC ACAGTCCGTT CGATCATCAT GATTCAAAAA GCACTTCTGA
 351  TTTCAGCGGC GGTGTAGACG GCGGTTTTAC TGTTTACCAA CTTCATCGAA
 401  CAGGGTCGGA AATCCATCCG GAGGATGGAT ATGACGGGCC GCAAGGCAGC
 451  GATTATCCGC CCCCCGGAGG AGCAAGGGAT ATATACAGCT ATTATGTCAA
 501  AGGAACTTCA ACAAAAACAA AGACTAATAT TGTCCCTCAA GCCCCATTTT
 551  CAGACCGTTG GCTAAAAGAA AATGCCGGTG CCGCCTCTGG TTTTTTCAGC
 601  CGTGCGGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA
 651  TTGGTGGGCT AACCGTATGG ATGATGTTCG CGGCATCGTC CAAGGTGCGG
 701  TTAATCCTTT TTTAATGGGT TTTCAAGGAG TAGGGATTGG GGCAATTACA
 751  GACAGTGCAG TAAGCCCGGT CACAGATACA GCCGCGCAGC AGACTCTACA
 801  AGGTATTAAT GATTTAGGAA AATTAAGTCC GGAAGCACAA CTTGCTGCCG
 851  CGAGCCTATT ACAGGACAGT GCTTTTGCGG TAAAAGACGG TATCAACTCT
 901  GCCAAACAAT GGGCTGATGC CCATCCAAAT ATAACAGCTA CTGCCCAAAC
 951  TGCCCTTTCC GCAGCAGAGG CCGCAGGTAC GGTTTGGAGA GGTAAAAAAG
1001  TAGAACTTAA CCCGACTAAA TGGGATTGGG TTAAAAATAC CGGTTATAAA
1051  AAACCTGCTG CCCGCCATAT GCAGACTTTA GATGGGGAGA TGGCAGGTGG
1101  GAATAAACCT ATTAAATCTT TACCAAACAG TGCCGCTGAA AAAAGAAAAC
1151  AAAATTTTGA GAAGTTTAAT AGTAACTGGA GTTCAGCAAG TTTTGATTCA
```

```
1201  GTGCACAAAA CACTAACTCC CAATGCACCT GGTATTTTAA GTCCTGATAA
1251  AGTTAAAACT CGATACACTA GTTTAGATGG AAAAATTACA ATTATAAAAG
1301  ATAACGAAAA CAACTATTTT AGAATCCATG ATAATTCACG AAAACAGTAT
1351  CTTGATTCAA ATGGTAATGC TGTGAAAACC GGTAATTTAC AAGGTAAGCA
1401  AGCAAAAGAT TATTTACAAC AACAAACTCA TATCAGGAAC TTAGACAAAT
1451  GA
```

This corresponds to the amino acid sequence <SEQ ID 166; ORF29-1>:

```
  1  MNLPIQKFMM LFAAAISLLQ IPISHANGLD ARLRDDMQAK HYEPGGKYHL
 51  FGNARGSVKK RVYAVQTFDA TAVSPVLPIT HERTGFEGVI GYETHFSGHG
101  HEVHSPFDHH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGS
151  DYPPPGGARD IYSYYVKGTS TKTKTNIVPQ APFSDRWLKE NAGAASGFFS
201  RADEAGKLIW ESDPNKNWWA NRMDDVRGIV QGAVNPFLMG FQGVGIGAIT
251  DSAVSPVTDT AAQQTLQGIN DLGKLSPEAQ LAAASLLQDS AFAVKDGINS
301  AKQWADAHPN ITATAQTALS AAEAAGTVWR GKKVELNPTK WDWVKNTGYK
351  KPAARHMQTL DGEMAGGNKP IKSLPNSAAE KRKQNFEKFN SNWSSASFDS
401  VHKTLTPNAP GILSPDKVKT RYTSLDGKIT IIKDNENNYF RIHDNSRKQY
451  LDSNGNAVKT GNLQGKQAKD YLQQQTHIRN LDK*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF29 shows 88.0% identity over a 125aa overlap with an ORF (ORF29a) from strain A of *N. meningitidis:*

```
                                                      10        20        30
orf29.pep                                   VSPVLPITHERTGFEGVIGYETHFSGHGHE
                                            |:|:||||||||||||:|||||||||||||
orf29a        EPGGKYHLFGNARGSVKNRVYAVQTFDATAVGPILPITHERTGFEGIIGYETHFSGHGHE
                      50        60        70        80        90       100

                      40        50        60        70        80        90
orf29.pep     VHSPFDHHDSKSTSDFSGGVDGGFTVYQLHRTWSEIHPEDEYDGPQAAXYPPPGGARDIY
              ||||||:||||||||||||||||||||||||| ||||||| |||||:: |||||||||||
orf29a        VHSPFDNHDSKSTSDFSGGVDGGFTVYQLHRTGSEIHPEDGYDGPQGSDYPPPGGARDIY
                     110       120       130       140       150       160

                     100       110       120
orf29.pep     SYYVKGTSTKTKTSIVPQAPFSDRWLEENAGAASG
                ||||||||||::|||:||||||||:||||||||
orf29a        XXYVKGTSTKTKSNIVPRAPFSDRWLKENAGAASGFFSRADEAGKLIWESDPNKNWWANR
                     170       180       190       200       210       220

orf29a        MDDIRGIVQGAVNPFLMGFQGVGIGAITDSAVSPVTDTAAQQTLQGXNHLGXLSPEAQLA
                     230       240       250       260       270       280
```

The complete length ORF29a nucleotide sequence <SEQ ID 167> is:

```
  1  ATGAATTNGC CTATTCAAAA ATTCATGATG CTGTTTGCAG CAGCAATATC
 51  GTNGCTGCAA ATCCCNATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC
101  GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGTAA ATACCATCTG
151  TTTGGTAATG CTCGCGGCAG TGTTAAAAAT CGGGTTTACG CCGTCCAAAC
201  ATTTGATGCA ACTGCGGTCG GCCCCATACT GCCTATTACA CACGAACGGA
251  CAGGATTTGA AGGCATTATC GGTTATGAAA CCCATTTTTC AGGACATGGA
301  CATGAAGTAC ACAGTCCGTT CGATAATCAT GATTCAAAAA GCACTTCTGA
351  TTTCAGCGGC GGCGTAGACG GTGGTTTTAC CGTTTACCAA CTTCATCGGA
401  CAGGGTCGGA AATCCATCCG GAGGATGGAT ATGACGGGCC GCAAGGCAGC
451  GATTATCCGC CCCCCGGAGG AGCAAGGGAT ATATACANNT ANTATGTCAA
501  AGGAACTTCA ACAAAAACAA AGAGTAATAT TGTTCCCCGA GCCCCATTTT
551  CAGACCGCTG GCTAAAAGAA AATGCCGGTG CCGCCTCTGG TTTTTTCAGC
601  CGTGCTGATG AAGCAGGAAA ACTGATATGG GAAAGCGACC CCAATAAAAA
651  TTGGTGGGCT AACCGTATGG ATGATATTCG CGGCATCGTC CAAGGTGCGG
701  TTAATCCTTT TTTAATGGGT TTTCAAGGAG TAGGGATTGG GGCAATTACA
751  GACAGTGCAG TAAGCCCGGT CACAGATACA GCCGCGCAGC AGACTCTACA
801  AGGTATNAAT CATTTAGGAA ANTTAAGTCC CGAAGCACAA CTTGCGGCTG
851  CAACCGCATT ACAAGACAGT GCTTTTGCGG TAAAAGACGG TATCAATTCC
901  GCCAGACAAT GGGCTGATGC CCATCCGAAT ATAACTGCAA CAGCCCAAAC
```

```
 951  TGCCCTTGCC GTAGCAGANG CCGCAACTAC GGTTTGGGGC GGTAAAAAAG
1001  TAGAACTTAA CCCGACCAAA TGGGATTGGG TTAAAAATAC NGGCTATAAN
1051  ACACCTGCTG TTCGCACCAT GCATACTTTG GATGGGGAAA TGGCCGGTGG
1101  GAATAGACCG CCTAAATCTA TAACGTCCAA CAGCAAAGCA GATGCTTCCA
1151  CACAACCGTC TTTACAAGCG CAACTAATTG GAGAACAAAT TANNNNNGGG
1201  CATGCTTATA ACAAGCATGT CATAAGACAA CAAGAATTTA CGGATTTAAA
1251  TATCAATTCA CCAGCAGATT TTGCTCGGCA TATTGAAAAT ATTGTTAGCC
1301  ATCCANCAAA TATGAAAGAG TTACCTCGCG GTAGAACTGC GTATTGGGAT
1351  NATAAAACAG GGACNATAGT TATCCGAGAT AAAAATTCTG ACGATGGAGG
1401  TACAGCATTT AGACCAACAT CAGGTAAAAA ATATTATGAT GATTTATAG
```

This encodes a protein having amino acid sequence <SEQ ID 168>:

```
  1  MNXPIQKFMM LFAAAISXLQ IPISHANGLD ARLRDDMQAK HYEPGGKYHL
 51  FGNARGSVKN RVYAVQTFDA TAVGPILPIT HERTGFEGII GYETHFSGHG
101  HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGS
151  DYPPPGGARD IYXXYVKGTS TKTKSNIVPR APFSDRWLKE NAGAASGFFS
201  RADEAGKLIW ESDPNKNWWA NRMDDIRGIV QGAVNPFLMG FQGVGIGAIT
251  DSAVSPVTDT AAQQTLQGXN HLGXLSPEAQ LAAATALQDS AFAVKDGINS
301  ARQWADAHPN ITATAQTALA VAXAATTVWG GKKVELNPTK WDWVKNTGYX
351  TPAVRTMHTL DGEMAGGNRP PKSITSNSKA DASTQPSLQA QLIGEQIXXG
401  HAYNKHVIRQ QEFTDLNINS PADFARHIEN IVSHPXNMKE LPRGRTAYWD
451  XKTGTIVIRD KNSDDGGTAF RPTSGKKYYD DL*
```

ORF29a and ORF29-1 show 90.1% identity in 385 aa overlap:

```
                     10        20        30        40        50        60
orf29a.pep   MNXPIQKFMMLFAAAISXLQIPISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKN
             || |||||||||||||| |||||||||||||||||||||||||||||||||||||||||:
orf29-1      MNLPIQKFMMLFAAAISLLQIPISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKK
                     10        20        30        40        50        60

                     70        80        90       100       110       120
orf29a.pep   RVYAVQTFDATAVGPILPITHERTGFEGIIGYETHFSGHGEVHSPFDNHDSKSTSDFSG
             |||||||||||||:|:||||||||||||:||||||||||||||||||:||||||||||||
orf29-1      RVYAVQTFDATAVSPVLPITHERTGFEGVIGYETHFSGHGEVHSPFDHHDSKSTSDFSG
                     70        80        90       100       110       120

                    130       140       150       160       170       180
orf29a.pep   GVDGGFTVYQLHRTGSEIHPEDGYDGPQGSDYPPPGGARDIYXXYVKGTSTKTKSNIVPR
             ||||||||||||||||||||||||||||||||||||||||||  ||||||||||:||||:
orf29-1      GVDGGFTVYQLHRTGSEIHPEDGYDGPQGSDYPPPGGARDIYSYYVKGTSTKTKTNIVPQ
                    130       140       150       160       170       180

                    190       200       210       220       230       240
orf29a.pep   APFSDRWLKENAGAASGFFSRADEAGKLIWESDPNKNWWANRMDDIRGIVQGAVNPFLMG
             |||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
orf29-1      APFSDRWLKENAGAASGFFSRADEAGKLIWESDPNKNWWANRMDDVRGIVQGAVNPFLMG
                    190       200       210       220       230       240

                    250       260       270       280       290       300
orf29a.pep   FQGVGIGAITDSAVSPVTDTAAQQTLQGXNHLGXLSPEAQLAAATALQDSAFAVKDGINS
             |||||||||||||||||||||||||||| | || ||||||||||: ||||||||||||||
orf29-1      FQGVGIGAITDSAVSPVTDTAAQQTLQGINDLGKLSPEAQLAAASLLQDSAFAVKDGINS
                    250       260       270       280       290       300

                    310       320       330       340       350       360
orf29a.pep   ARQWADAHPNITATAQTALAVAXAATTVWGGKKVELNPTKWDWVKNTGYXTPAVRTMHTL
             |:|||||||||||||||||::| || ||| |||||||||||||||||||  ||:| |:||
orf29-1      AKQWADAHPNITATAQTALSAAEAAGTVWRGKKVELNPTKWDWVKNTGYKKPAARHMQTL
                    310       320       330       340       350       360

                    370       380       390       400       410       420
orf29a.pep   DGEMAGGNRPPKSITSNSKADASTQPSLQAQLIGEQIXXGHAYNKHVIRQQEFTDLNINS
             ||||||||:| ||:  || |:   |
orf29-1      DGEMAGGNKPIKSLP-NSAAEKRKQNFEKFNSNWSSASFDSVHKTLTPNAPGILSPDKVK
                    370       380       390       400       410
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF29 shows 88.8% identity over a 125aa overlap with a predicted ORF (ORF29.ng) from *N. gonorrhoeae:*

```
orf29.pep                                           VSPVLPITHERTGFEGVIGYETHFSGHGHE   30
                                                    |:|:||||||||||||||||||||||||||
orf29ng      EPGGKYHLFGNARGSVKNRVCAVQTFDATAVGPILPITHERTGFEGVIGYETHFSGHGHE  102

orf29.pep    VHSPFDHHDSKSTSDFSGGVDGGFTVYQLHRTWSEIHPEDEYDGPQAAXYPPPGGARDIY   90
             ||||||:||||||||||||||||||||||||| ||||||| |||||:: ||||||||||
orf29ng      VHSPFDNHDSKSTSDFSGGVDGGFTVYQLHRTGSEIHPEDGYDGPQGGGYPPPGGARDIY  162

orf29.pep    SYYVKGTSTKTKTSIVPQAPFSDRWLEENAGAASG                          125
             ||::|||||||| : |||||||||||:||||||||
orf29ng      SYHIKGTSTKTKINTVPQAPFSDRWLKENAGAASGFLSRADEAGKLIWENDPDKNWRANR  222
```

The complete length ORF29ng nucleotide sequence <SEQ ID 169> is predicted to encode a protein having amino acid sequence <SEQ ID 170>:

```
  1  MNLPIQKFMM LFAAAISLLQ IPISHANGLD ARLRDDMQAK HYEPGGKYHL
 51  FGNARGSVKN RVCAVQTFDA TAVGPILPIT HERTGFEGVI GYETHFSGHG
101  HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP EDGYDGPQGG
151  GYPPPGGARD IYSYHIKGTS TKTKINTVPQ APFSDRWLKE NAGAASGFLS
201  RADEAGKLIW ENDPDKNWRA NRMDDIRGIV QGAVNPFLTG FQGLGVGAIT
251  DSAVSPVTYA AARKTLQGIH NLGNLSPEAQ LAAATALQDS AFAVKDSINS
301  ARQWADAHPN ITATAQTALA VTEAATTVWG GKKVELNPAK WDWVKNTGYK
351  KPAARHMQTV DGEMAGGNKP LESKNTVTTN NFFENTGYTE KVLRQASNGD
401  YHGFPQSVDA FSENGTVIQI VGGDNIVRHK LYIPGSYKGK DGNFEYIREA
451  DGKINHRLFV PNQQLPEK*
```

In a second experiment, the following DNA sequence <SEQ ID 171> was identified:

```
   1  atgAATTTGC CTATTCAAAA ATTCATGATG ctgttggcAg cggcaatatc
  51  gatgctGCat ATCCCCATTA GTCATGCGAA CGGTTTGGAT GCCCGTTTGC
 101  GCGATGATAT GCAGGCAAAA CACTACGAAC CGGGTGGCAA ATACCATCTG
 151  TTTGGTAATG CTCGCGGCAG TGTTAAAAAT CGGGTTTGCG CCGTCCAAAC
 201  ATTTGATGCA ACTGCGGTCG GCCCCATACT GCCTATTACA CACGAACGGA
 251  CAGGATTTGA AGGTGTTATC GGCTATGAAA CCCATTTTTC AGGACACGGA
 301  CACGAAGTAC ACAGTCCGTT CGATAATCAT GATTCAAAAA GCACTTCTGA
 351  TTTCAGCGGC GGCGTAGACG GCGGTTTTAC CGTTTACCAA CTTCATCGGA
 401  CAGGGTCGGA AATACATCCC GCAGACGGAT ATGACGGGCC TCAAGGCGGC
 451  GGTTATCCGG AACCACAAGG GGCAAGGGAT ATATACAGCT ACCATATCAA
 501  AGGAACTTCA ACCAAAACAA AGATAAACAC TGTTCCGCAA GCCCCTTTTT
 551  CAGACCGCTG GCTAAAAGAA AATGCCGGTG CCGCTTCCGG TTTTCTCAGC
 601  CGTGCGGATG AAGCAGGAAA ACTGATATGG GAAAACGACC CCGATAAAAA
 651  TTGGCGGGCT AACCGTATGG ATGATATTCG CGGCATCGTC CAAGGTGCGG
 701  TTAATCCTTT TTTAACGGGT TTTCAAGGGG TAGGGATTGG GGCAATTACA
 751  GACAGTGCGG TAAGCCCGGT CACAGATACA GCCGCTCAGC AGACTCTACA
 801  AGGTATTAAT GATTTAGGAA ATTTAAGTCC GGAAGCACAA CTTGCCGCCG
 851  CGAGCCTATT ACAGGACAGT GCCTTTGCGG TAAAAGACGG CATCAATTCC
 901  GCCAGACAAT GGGCTGATGC CCATCCGAAT ATAACAGCAA CAGCCCAAAC
 951  TGCCCTTGCC GTAGCAGAGG CCGCAGGTAC GGTTTGGCGC GGTAAAAAAG
1001  TAGAACTTAA CCCGACCAAA TGGGATTGGG TTAAAAATAC CGGCTATAAA
1051  AAACCTGCTG CCCGCCATAT GCAGACTGTA GATGGGGAGA TGGCAGGGGG
1101  GAATAGACCG CCTAAATCTA TAACGTCGGA AGGAAAAGCT AATGCTGCAA
1151  CCTATCCTAA GTTGGTTAAT CAGCTAAATG AGCAAAACTT AAATAACATT
1201  GCGGCTCAAG ATCCAAGATT GAGTCTAGCT ATTCATGAGG GTAAAAAAAA
1251  TTTTCCAATA GGAACTGCAA CTTATGAAGA GGCAGATAGA CTAGGTAAAA
1301  TTTGGGTTGG TGAGGGTGCA AGACAAACTA GTGGAGGCGG ATGGTTAAGT
1351  AGAGATGGCA CTCGACAATA TCGGCCACCA ACAGAAAAAA AATCACAATT
1401  TGCAACTACA GGTATTCAAG CAAATTTTGA AACTTATACT ATTGATTCAA
1451  ATGAAAAAAG AAATAAAATT AAAAATGGAC ATTTAAATAT TAGGTAA
```

This encodes a protein having amino acid sequence <SEQ ID 172; ORF29ng-1>:

```
  1  MNLPIQKFMM LLAAAISMLH IPISHANGLD ARLRDDMQAK HYEPGGKYHL
 51  FGNARGSVKN RVCAVQTFDA TAVGPILPIT HERTGFEGVI GYETHFSGHG
```

```
101  HEVHSPFDNH DSKSTSDFSG GVDGGFTVYQ LHRTGSEIHP ADGYDGPQGG
151  GYPEPQGARD IYSYHIKGTS TKTKINTVPQ APFSDRWLKE NAGAASGFLS
201  RADEAGKLIW ENDPDKNWRA NRMDDIRGIV QGAVNPFLTG FQGVGIGAIT
251  DSAVSPVTDT AAQQTLQGIN DLGNLSPEAQ LAAASLLQDS AFAVKDGINS
301  ARQWADAHPN ITATAQTALA VAEAAGTVWR GKKVELNPTK WDWVKNTGYK
351  KPAARHMQTV DGEMAGGNRP PKSITSEGKA NAATYPKLVN QLNEQNLNNI
401  AAQDPRLSLA IHEGKKNFPI GTATYEEADR LGKIWVGEGA RQTSGGGWLS
451  RDGTRQYRPP TEKKSQFATT GIQANFETYT IDSNEKRNKI KNGHLNIR*
```

ORF29ng-1 and ORF29-1 show 86.0% identity in 401 aa overlap:

```
                        10        20        30        40        50        60
orf29ng-1.pep   MNLPIQKFMMLLAAAISMLHIPISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKN
                |||||||||||:|||||:|:|||||||||||||||||||||||||||||||||||||||:
orf29-1         MNLPIQKFMMLFAAAISLLQIPISHANGLDARLRDDMQAKHYEPGGKYHLFGNARGSVKK
                        10        20        30        40        50        60

                        70        80        90       100       110       120
orf29ng-1.pep   RVCAVQTFDATAVGPILPITHERTGFEGVIGYETHFSGHGHEVHSPFDNHDSKSTSDFSG
                || ||||||||||:|:|||||||||||||||||||||||||||||||||:||||||||||
orf29-1         RVYAVQTFDATAVSPVLPITHERTGFEGVIGYETHFSGHGHEVHSPFDHHDSKSTSDFSG
                        70        80        90       100       110       120

                       130       140       150       160       170       180
orf29ng-1.pep   GVDGGFTVYQLHRTGSEIHPADGYDGPQGGGYPEPQGARDIYSYHIKGTSTKTKINTVPQ
                |||||||||||||||||||| ||||||||: || | ||||||||::|||||||| | |||
orf29-1         GVDGGFTVYQLHRTGSEIHPEDGYDGPQGSDYPPPGGARDIYSYYVKGTSTKTKTNIVPQ
                       130       140       150       160       170       180

                       190       200       210       220       230       240
orf29ng-1.pep   APFSDRWLKENAGAASGFLSRADEAGKLIWENDPDKNWRANRMDDIRGIVQGAVNPFLTG
                ||||||||||||||||||:||||||||||||:||:|||| |||||:|||||||||||| |
orf29-1         APFSDRWLKENAGAASGFFSRADEAGKLIWESDPNKNWWANRMDDVRGIVQGAVNPFLMG
                       190       200       210       220       230       240

                       250       260       270       280       290       300
orf29ng-1.pep   FQGVGIGAITDSAVSPVTDTAAQQTLQGINDLGNLSPEAQLAAASLLQDSAFAVKDGINS
                |||||||||||||||||||||||||||||||||:||||||||||||||||||||||||||
orf29-1         FQGVGIGAITDSAVSPVTDTAAQQTLQGINDLGKLSPEAQLAAASLLQDSAFAVKDGINS
                       250       260       270       280       290       300

                       310       320       330       340       350       360
orf29ng-1.pep   ARQWADAHPNITATAQTALAVAEAAGTVWRGKKVELNPTKWDWVKNTGYKKPAARHMQTV
                |:|||||||||||||||||::||||||||||||||||||||||||||||||||||||||:
orf29-1         AKQWADAHPNITATAQTALSAAEAAGTVWRGKKVELNPTKWDWVKNTGYKKPAARHMQTL
                       310       320       330       340       350       360

                       370       380       390       400       410      419
orf29ng-1.pep   DGEMAGGNRPPKSI-TSEGKANAATYPKLVNQLNEQNLNNIAAQDPRLSLAIHEGKKNFP
                ||||||||:| ||: :| ::     :: |: :: :  ::::
orf29-1         DGEMAGGNKPIKSLPNSAAEKRKQNFEKFNSNWSSASFDSVHKTLTPNAPGILSPDKVKT
                       370       380       390       400       410       420

                     420       430       440       450       460       470      479
orf29ng-1.pep   IGTATYEEADRLGKIWVGEGARQTSGGGWLSRDGTRQYRPPTEKKSQFATTGIQANFETY

orf29-1         RYTSLDGKITIIKDNENNYFRIHDNSRKQYLDSNGNAVKTGNLQGKQAKDYLQQQTHIRN
                       430       440       450       460       470       480
```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 21

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 173>:

```
  1  ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
 51  CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAATGTTCC
101  ACACGCGGGC AGATGCACCG ATGCAG...
```

This corresponds to the amino acid sequence <SEQ ID 174; ORF30>:

```
  1  MKKQITAAVM MLSMIAPAMA NGLDNQAFED QMFHTRADAP MQ..
```

Further work revealed the complete nucleotide sequence <SEQ ID 175>:

```
  1  ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
 51  CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTCC
101  ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG
151  ATGAAGGAGA CAGAGGGGGC GTTTCTTCCA TTGGCTATCT TGGGTGGTGC
201  TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
251  GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
301  CCTGGTGGTG TAGGCGCCGC AGGAAAGGTT GTTTCCTTTG CTAAATATGG
351  ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCCT TTCGGTAATA
401  GAACAGGTCA TCCTATTGGA AAATTTCCCC ATTATCATCG TCGAGTTACG
451  GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
501  TTGGGAATCA AAATCTACGG ACAGATCATG GAAAAACCGC TTCTAA
```

This corresponds to the amino acid sequence <SEQ ID 176; ORF30-1>:

```
  1  MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
 51  MKETEGAFLP LAILGGAAIG MWTQHGFSYA TTGRPASVRD VAIAGGLGAI
101  PGGVGAAGKV VSFAKYGREI KIGNNMRIAP FGNRTGHPIG KFPHYHRRVT
151  DNTGKTLPGQ GIGRHRPWES KSTDRSWKNR F*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF30 shows 97.6% identity over a 42aa overlap with an ORF (ORF30a) from strain A of *N. meningitidis:*

```
                  10        20        30        40
orf30.pep   MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ
            ||||||||||||||||||||||||||||||:|||||||||||
orf30a      MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP
                  10        20        30        40        50        60

orf30a      LXILGGAAIGMWTQHGFSYATTGRPASVRDVAIAGGLGAIPGXVGAAGKVVSFAKYGREI
                  70        80        90       100       110       120
```

The complete length ORF30a nucleotide sequence <SEQ ID 177> is:

```
  1  ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATTGCCCC
 51  CGCAATGGCA AACGGCTTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTCC
101  ACACGCGGGC AGATGCACCG ATGCAGTTGG CGGAGCTTTC TCAAAAGGAG
151  ATGAAGGANA CAGNGGGGGC GTTTCTTCCA TTGGNTATCT TGGGTGGTGC
201  TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
251  GACCAGCTTC TGTTAGAGAT GTTGCTATTG CTGGCGGATT AGGCGCAATT
301  CCTGGTGNTG TAGGCGCCGC AGGAAAGGTT GTTTCCTTTG CTAAATATGG
351  ACGTGAGATT AAAATCGGCA ATAATATGCG GATAGCCCCT TTCGGTAATA
401  GAACAGGTCA TCCTATTGGN AAATTTCCCC ATTATCATCG TCGAGTTACG
451  GATAATACGG GCAAGACTTT GCCTGGACAG GGAATTGGTC GTCATCGCCC
501  TTGGGAATCA AAATCTACGG ACAGATCATG GAAAAACCGC TTCTAA
```

This encodes a protein having amino acid sequence <SEQ ID 178>:

```
  1  MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
 51  MKXTXGAFLP LXILGGAAIG MWTQHGFSYA TTGRPASVRD VAIAGGLGAI
101  PGXVGAAGKV VSFAKYGREI KIGNNMRIAP FGNRTGHPIG KFPHYHRRVT
151  DNTGKTLPGQ GIGRHRPWES KSTDRSWKNR F*
```

ORF30a and ORF30-1 show 97.8% identity in 181 aa overlap:

```
orf30a.pep   MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKXTXGAFLP  60
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||| | |||||
orf30-1          MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP  60

orf30a.pep       LXILGGAAIGMWTQHGFSYATTGRPASVRDVAIAGGLGAIPGXVGAAGKVVSFAKYGREI 120
                 | |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
orf30-1          LAILGGAAIGMWTQHGFSYATTGRPASVRDVAIAGGLGAIPGGVGAAGKVVSFAKYGREI 120

orf30a.pep       KIGNNMRIAPFGNRTGHPIGKFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf30-1          KIGNNMRIAPFGNRTGHPIGKFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR 180

orf30a.pep       FX
                 ||
orf30-1          FX
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF30 shows 97.6% identity over a 42aa overlap with a predicted ORF (ORF30.ng) from *N. gonorrhoeae:*

```
orf30.pep      MKKQITAAVMMLSMIAPAMANGLDNQAFEDQMFHTRADAPMQ                    42
               |||||||||||||||||||||||||||||||:||||||||||
orf30ng        MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP  60
```

The complete length ORF30ng nucleotide sequence <SEQ ID 179> is

```
  1  ATGAAAAAAC AAATCACCGC AGCCGTAATG ATGCTGTCTA TGATCGCCCC
 51  CGCAATGGCA AACGGATTGG ACAATCAGGC ATTTGAAGAC CAAGTGTTCC
101  ACACGCGGGC AGATGCGCCG ATGCAGTTGG CGGAGCTTTC TCAGAAGGAG
151  ATGAAGGAGA CTGAAGGGGC TTTTCTTCCA TTGGCTATCT TGGGTGGTGC
201  TGCCATTGGT ATGTGGACAC AGCATGGTTT TAGTTATGCA ACGACAGGCA
251  GACCAGCTTC TGTTAGAGAT GTTGCTGGCG GATTAGGCGC AATTCCTGGT
301  GATGTAGGTG CTGCAGGAAA GGTTGTTTCC TTTGCTAAAT ATGGACGTGA
351  GATTAAAATC GGCAATAATA TGCGGATAGC CCCTTTCGGT AATAGAACAG
401  GTCATCCTAT TGGAAAATTT CCCCATTATC ATCGTCGAGT TACGGATAAT
451  ACGGGCAAGA CTTTGCCTGG ACAGGGAATT GGTCGTCATC GCCCTTGGGA
501  ATCAAAATCT ACGGACAGAT CATGGAAAAA CCGCTTCTAA
```

This encodes a protein having amino acid sequence <SEQ ID 180>:

```
  1  MKKQITAAVM MLSMIAPAMA NGLDNQAFED QVFHTRADAP MQLAELSQKE
 51  MKETEGAFLP LAILGGAAIG MWTQHGFSYA TTGRPASVRD VAGGLGAIPG
101  DVGAAGKVVS FAKYGREIKI GNNMRIAPFG NRTGHPIGKF PHYHRRVTDN
151  TGKTLPGQGI GRHRPWESKS TDRSWKNRF*
```

ORF30ng and ORF30-1 show 98.3% identity in 181 aa overlap:

```
                          10        20        30        40        50        60
orf30ng.pep      MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf30-1          MKKQITAAVMMLSMIAPAMANGLDNQAFEDQVFHTRADAPMQLAELSQKEMKETEGAFLP
                          10        20        30        40        50        60

                          70        80        90         100       110
orf30ng.pep      LAILGGAAIGMWTQHGFSYATTGRPASVRDVA--GGLGAIPGDVGAAGKVVSFAKYGREI
                 ||||||||||||||||||||||||||||||||  |||||||| |||||||||||||||||
orf30-1          LAILGGAAIGMWTQHGFSYATTGRPASVRDVAIAGGLGAIPGGVGAAGKVVSFAKYGREI
                          70        80        90        100       110       120

                120       130       140       150       160       170
orf30ng.pep      KIGNNMRIAPFGNRTGHPIGKFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf30-1          KIGNNMRIAPFGNRTGHPIGKFPHYHRRVTDNTGKTLPGQGIGRHRPWESKSTDRSWKNR
                         130       140       150       160       170       180

                180
orf30ng.pep      FX
                 ||
orf30-1          FX
```

Based on this analysis, including the presence of a putative leader sequence in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 22

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 181>:

```
  1  ATGAATAAAA CTCTCTATCG TGTAATTTTC AACCGCAAAC GTGGGGCTGT
 51  GrTAGCCGTT GCTGAAACTA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA
101  GTGATTCAGG CAGCGCTCAT GTGAAATCTG TTCCTTTTGG TACTACTCAT
151  GCACCTGTTT GTg.CGTTaC AAATATCTTT TCTTTTTCTT TATTGGGCTT
201  TTCTTTATGT TTGGCTGTAG GtacGGyCAA TATTGCTTTT GCTGATGGCA
251  TT..
```

This corresponds to the amino acid sequence <SEQ ID 182; ORF31>:

```
  1  MNKTLYRVIF NRKRGAVXAV AETTKREGKS CADSDSGSAH VKSVPFGTTH
 51  APVCXVTNIF SFSLLGFSLC LAVGTXNIAF ADGI..
```

Further work revealed a further partial nucleotide sequence <SEQ ID 183>:

```
  1  ATGAATAAAA CTCTCTATCG TGTAATTTTC AACCGCAAAC GTGGGGCTGT
 51  GGTAGCCGTT GCTGAAACTA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA
101  GTGATTCAGG CAGCGCTCAT GTGAAATCTG TTCCTTTTGG TACTACTCAT
151  GCACCTGTTT GTCGTTCAAA TATCTTTTCT TTTTCTTTAT TGGGCTTTTC
201  TTTATGTTTG GCTGTAGGTA CGGCCAATAT TGCTTTTGCT GATGGCATT..
```

This corresponds to the amino acid sequence <SEQ ID 184; ORF31-1>:

```
  1  MNKTLYRVIF NRKRGAVVAV AETTKREGKS CADSDSGSAH VKSVPFGTTH
 51  APVCRSNIFS FSLLGFSLCL AVGTANIAFA DGI..
```

Computer analysis of this amino acid sequence gave the following results:

<u>Homology with a predicted ORF from *N.gonorrhoeae*</u>

ORF31 shows 76.2% identity over a 84aa overlap with a predicted ORF (ORF31.ng) from *N. gonorrhoeae:*

```
orf31.pep   MNKTLYRVIFNRKRGAVXAVAETTKREGKSCADSDSGSAHVKSVPFGTTHAPVCXVTNIF   60
            ||||||||||||||||| |||||||||||||||| |||::|||| |  ||      :: |
orf31ng     MNKTLYRVIFNRKRGAVVAVAETTKREGKSCADSGSGSVYVKSVSFIPTH------SKAF   54

orf31.pep   SFSLLGFSLCLAVGTXNIAFADGI                                       84
             || ||||||||:|| ||||||||
orf31ng     CFSALGFSLCLALGTVNIAFADGIITDKAAPKTQQATILQTGNGIPQVNIQTPTSAGVSV  114
```

The complete length ORF31ng nucleotide sequence <SEQ ID 185> is:

```
  1  ATGAACAAAA CCCTCTATCG TGTGATTTTC AACCGCAAAC GCGGTGCTGT
 51  GGTAGCTGTT GCCGAAACCA CCAAGCGCGA AGGTAAAAGC TGTGCCGATA
101  GTGGTTCGGG CAGCGTTTAT GTGAAATCCG TTTCTTTCAT TCCTACTCAT
151  TCCAAAGCCT TTTGTTTTTC TGCATTAGGC TTTTCTTTAT GTTTGGCTTT
201  GGGTACGGTC AATATTGCTT TTGCTGACGG CATTATTACT GATAAAGCTG
251  CTCCTAAAAC CCAACAAGCC ACGATTCTGC AAACAGGTaa cGGCATACCG
301  CAAGTCAATA TTCAAACCCC TACTTCGGCA GGGGTTTCTG TTAATCAATA
351  TGCCCAGTTT GATGTGGGTA ATCGCGGGGC GATTTTAAAC AACAGTCGCA
401  GCAACACCCA AACACAGCTA GGCGGTTGGA TTCAAGGCAA TCCTTGGTTG
451  ACAAGGGGCG AAGCACGTGT GGTTGTAAAC CAAATCAACA GCAGCCATCC
501  TTCACAACTG AATGGCTATA TTGAAGTGGG TGGACGACGT GCAGAAGTCG
551  TTATTGCCAA TCCGGCAGGG ATTGCAGTCA ATGGTGGTGG TTTTATCAAT
601  GCTTCCCGTG CCACTTTGAC GACAGGCCAA CCGCAATATC AAGCAGGAGA
651  CTTTAGCGGC TTTAAGATAA GGCAAGGCAA TGCTGTAATC GCCGGACACG
```

```
701  GTTTGGATGC CCGTGATACC GATTTCACAC GTATTCTTGT ATGCCAACAA
751  AATCACCTTG ATCAGTACGG CCGAACAAGC AGGCATTCGT AA
```

This encodes a protein having amino acid sequence <SEQ ID 186>:

```
  1  MNKTLYRVIF NRKRGAVVAV AETTKREGKS CADSGSGSVY VKSVSFIPTH
 51  SKAFCFSALG FSLCLALGTV NIAFADGIIT DKAAPKTQQA TILQTGNGIP
101  QVNIQTPTSA GVSVNQYAQF DVGNRGAILN NSRSNTQTQL GGWIQGNPWL
151  TRGEARVVVN QINSSHPSQL NGYIEVGGRR AEVVIANPAG IAVNGGGFIN
201  ASRATLTTGQ PQYQAGDFSG FKIRQGNAVI AGHGLDARDT DFTRILVCQQ
251  NHLDQYGRTS RHS*
```

This gonococcal protein shares 50% identity over a 149aa overlap with the pore-forming hemolysins-like HecA protein from *Erwinia chrysanthemi* (accession number L39897):

```
orf31ng  96  GNGIPQVNIQTPTSAGVSVNQYAQFDVGNRGAILNNSRSN-TQTQLGGWIQGNPWLTRGE 154
             GNG+P VNI TP ++G+S N+Y  F+V NRG ILNN  +  T +QLGG IQ NP L
HecA     45  GNGVPVVNIATPDASGLSHNRYHDFNVDNRGLILNNGTARLTPSQLGGLIQNNPNLNGRA 104

Orf31ng  155 ARVVVNQINSSHPSQLNGYIEVGGRRAEVVIANPAGIAVNGGGFINASRATLTTGQPQYQ 214
             A  ++N++ S + S+L GY+EV G+ A VV+ANP GI  +G GF+N  R TLTTG PQ+
HecA     105 AAAILNEVVSPNRSRLAGYLEVAGQAANVVVANPYGITCSGCGFLNTPRLTLTTGTPQFD 164

Orf31ng  215 -AGDFSGFKIRQGNAVIAGHGLDARDTDF 242
              AG  SG  +R G+ +I G GLDA  +D+
HecA     165 AAGGLSGLDVRGGDILIDGAGLDASRSDY 193
```

Furthermore, ORF31ng and ORF31-1 show 79.5% identity in 83 aa overlap:

```
                      10        20        30        40        50        60
orf31-1.pep   MNKTLYRVIFNRKRGAVVAVAETTKREGKSCADSDSGSAHVKSVPFGTTHAPVCRSNIFS
              |||||||||||||||||||||||||||||||||| |||::|||| |  ||     |: |
orf31ng       MNKTLYRVIFNRKRGAVVAVAETTKREGKSCADSGSGSVYVKSVSFIPTH-----SKAFC
                      10        20        30        40        50

                      70        80
orf31-1.pep   FSLLGFSLCLAVGTANIAFADGI
              || ||||||||:||:||||||||
orf31ng       FSALGFSLCLALGTVNIAFADGIITDKAAPKTQQATILQTGNGIPQVNIQTPTSAGVSVN
                 60        70        80        90       100       110
```

On this basis, including the homology with hemolysins, and also with adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 23

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 187>:

```
  1  ATGAATACTC CTCCTTTTGT CTGTTGGATT TTTTGCAAGG TCATCGACAA
 51  TTTCGGCGAC ATCGGCGTTT CGTGGCGGCT CGCCCGTGTT TTGCACCGCG
101  AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT
151  GCGCTTTGCC CTGATTTGCC CGATGTTCCC TGCGTTCATC AGGATATTCA
201  TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCG..
```

This corresponds to the amino acid sequence <SEQ ID 188; ORF32>:

```
 1  MNTPPFVCWI FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR
51  ALCPDLPDVP CVHQDIHVRT WHSDAADIDT A..
```

Further work revealed the complete nucleotide sequence <SEQ ID 189>:

```
  1  ATGAATACTC CTCCTTTTGT CTGTTGGATT TTTTGCAAGG TCATCGACAA
 51  TTTCGGCGAC ATCGGCGTTT CGTGGCGGCT CGCCCGTGTT TTGCACCGCG
101  AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT
```

```
 151  GCGCTTTGCC CTGATTTGCC CGATGTTCCC TGCGTTCATC AGGATATTCA
 201  TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTTC
 251  CCGATGTCGT CATCGAAACT TTTGCCTGCG ACCTGCCCGA AAATGTGCTG
 301  CACATTATCC GCCGACACAA GCCGCTTTGG CTGAATTGGG AATATTTGAG
 351  CGCGGAGGAA AGCAATGAAA GGCTGCATCT GATGCCTTCG CCGCAGGAGG
 401  GTGTTCAAAA ATATTTTTGG TTTATGGGTT TCAGCGAAAA AAGCGGCGGG
 451  TTGATACGCG AACGTGATTA CTGCGAAGCC GTCCGTTTCG ATACTGAAGC
 501  CCTGCGAGAG CGGCTGATGC TGCCCGAAAA AAACGCCTCC GAATGGCTGC
 551  TTTTCGGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA
 601  CAGGCAGGCA GCCCGATGAC ACTGTTGCTG GCGGGGACGC AAATCATCGA
 651  CAGCCTCAAA CAAAGCGGCG TTATTCCGCA AGATGCCCTG CAAAACGACG
 701  GCGATGTTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCCTTTCGTG
 751  CCGCAACAGG ACTTCGACCA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT
 801  CCGCGGCGAA GACAGTTTCG TGCGCGCCCA GCTTGCGGGC AAACCCTTCT
 851  TTTGGCACAT CTACCCGCAA GACGAGAATG TCCATCTCGA CAAACTCCAC
 901  GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCCGAAA CCGTGTCGGC
 951  ACACCGCCGT CTTTCGGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA
1001  CACAACGCCT CGAATGTTGG CAAACCCTGC AACAACATCA AAACGGCTGG
1051  CGGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTCGGGC AGCCGTCAGC
1101  TCCTGAAAAA CTCGCTGCCT TTGTTTCAAA GCATCAAAAA ATACGCTAG
```

This corresponds to the amino acid sequence <SEQ ID 190; ORF32-1>:

```
  1  MNTPPFVCWI FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR
 51  ALCPDLPDVP CVHQDIHVRT WHSDAADIDT APVPDVVIET FACDLPENVL
101  HIIRRHKPLW LNWEYLSAEE SNERLHLMPS PQEGVQKYFW FMGFSEKSGG
151  LIRERDYCEA VRFDTEALRE RLMLPEKNAS EWLLFGYRSD VWAKWLEMWR
201  QAGSPMTLLL AGTQIIDSLK QSGVIPQDAL QNDGDVFQTA SVRLVKIPFV
251  PQQDFDQLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPQ DENVHLDKLH
301  AFWDKAHGFY TPETVSAHRR LSDDLNGGEA LSATQRLECW QTLQQHQNGW
351  RQGAEDWSRY LFGQPSAPEK LAAFVSKHQK IR*w
```

Computer analysis of this amino acid sequence gave the following results:

<u>Homology with a predicted ORF from *N.meningitidis* (strain A)</u>

ORF32 shows 93.8% identity over a 81aa overlap with an ORF (ORF32a) from strain A of *N. meningitidis:*

```
                     10        20        30        40        50        60
orf32.pep   MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVP
            ||||||    ||||||||||||||||||||||||||||||||||||||||||||||||||
orf32a      MNTPPFSAGXFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVX
                     10        20        30        40        50        60

                     70        80
orf32.pep   CVHQDIHVRTWHSDAADIDTA
            |||||||||||||||||||||
orf32a      CVHQDIHVRTWHSDAADIDTAPVXDVVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEX
                     70        80        90       100       110       120
```

The complete length ORF32a nucleotide sequence <SEQ ID 191> is:

```
   1  ATGAATACTC CTCCTTTTTC TGCTGGANTT TTTTGCAAGG TCATCGACAA
  51  TTTCGGCGAC ATCGGCGTTT CGTGGCGGCT TGCCCGTGTT TTGCACCGCG
 101  AACTCGGTTG GCAGGTGCAT TTGTGGACGG ACGATGTGTC CGCCTTGCGT
 151  GCGCTTTGCC CTGATTTGCC CGATGTTCNC TGCGTTCATC AGGATATTCA
 201  TGTCCGCACT TGGCATTCCG ATGCGGCAGA TATTGATACC GCGCCTGTTC
 251  NCGATGTCGT CATCGAAACT TTTGCCTGCG ACCTGCCCGA AAATGTGCTG
 301  CACATCATCC GCCGACACAA GCCGCTTTGG CTGAANTGGG AATATTTGAG
 351  CGCGGAGGAN AGCAATGAAA GGCTGCACNT GATGCCTTCG CCGCAGGAGA
 401  GTGTTCNAAA ATANTTTTGG TTTATGGGTT TCAGCGAANN NAGCGGCGGA
 451  CTGATACGCG AACGCGATTA CTGCGAAGCC GTCCGTTTCG ATAGCGGAGC
 501  CTTGCGCAAG AGGCTGATGC TTCCCGAAAA AAACGNCCCC GAATGGCTGC
 551  TTTTCGGCTA TCGGAGCGAT GTTTGGGCAA AGTGGCTGGA AATGTGGCGA
 601  CAGGCAGGCA GTCCGTTGAC ACTTTTGCTG GCNGGGGCGC ANATTATCGA
 651  CAGCCTCAAA CAAAACGGCG TTATTCCGCA AGATGCCCTG CAAAACGACG
 701  GCGATGTTTT TCAGACGGCA TCCGTCCGCC TCGTCAAAAT CCCTTTCGTG
 751  CCGCAACAGG ACTTCGACAA ACTGCTGCAC CTTGCCGACT GCGCCGTCAT
```

```
 801  CCGCGGCGAA GACAGTTTCG TGCGCGCCCA GCTTGCGGGC AAACCCTTCT
 851  TTTGGCACAT CTACCCGCAA GATGAGAATG TCCATCTCGA CAAACTCCAC
 901  GCCTTTTGGG ATAAGGCACA CGGTTTCTAC ACGCCCGAAA CCGCATCGGC
 951  ACACCGCCGC CTTTCAGACG ACCTCAACGG CGGAGAGGCT TTATCCGCAA
1001  CACAACGCCT CGAATGTTGG CAAATCCTGC AACAACATCA AAACGGCTGG
1051  CGGCAAGGCG CGGAGGATTG GAGCCGTTAT CTTTTTGGGC AGCCTTCCGC
1101  ATCCGAAAAA CTCGCCGCCT TTGTTTCAAA GCATCAAAAA ATACGCTAG
```

This encodes a protein having amino acid sequence <SEQ ID 192>:

```
  1  MNTPPFSAGX FCKVIDNFGD IGVSWRLARV LHRELGWQVH LWTDDVSALR
 51  ALCPDLPDVX CVHQDIHVRT WHSDAADIDT APVXDVVIET FACDLPENVL
101  HIIRRHKPLW LXWEYLSAEX SNERLHXMPS PQESVXKXFW FMGFSEXSGG
151  LIRERDYCEA VRFDSGALRK RLMLPEKNXP EWLLFGYRSD VWAKWLEMWR
201  QAGSPLTLLL AGAXIIDSLK QNGVIPQDAL QNDGDVFQTA SVRLVKIPFV
251  PQQDFDKLLH LADCAVIRGE DSFVRAQLAG KPFFWHIYPQ DENVHLDKLH
301  AFWDKAHGFY TPETASAHRR LSDDLNGGEA LSATQRLECW QILQQHQNGW
351  RQGAEDWSRY LFGQPSASEK LAAFVSKHQK IR*
```

ORF32a and ORF32-1 show 93.2% identity in 382 aa overlap:

```
                       10        20        30        40        50        60
orf32-1.pep    MNTPPFVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVP
               ||||||    ||||||||||||||||||||||||||||||||||||||||||||||||| 
orf32a         MNTPPFSAGXFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDVX
                       10        20        30        40        50        60

                       70        80        90       100       110       120
orf32-1.pep    CVHQDIHVRTWHSDAADIDTAPVPDVVIETFACDLPENVLHIIRRHKPLWLNWEYLSAEE
               ||||||||||||||||||||||| ||||||||||||||||||||||||||| ||||||| 
orf32a         CVHQDIHVRTWHSDAADIDTAPVXDVVIETFACDLPENVLHIIRRHKPLWLXWEYLSAEX
                       70        80        90       100       110       120

                      130       140       150       160       170       180
orf32-1.pep    SNERLHLMPSPQEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDTEALRERLMLPEKNAS
               |||||| ||||||:| | |||||||| |||||||||||||||||: |||:||||||||  
orf32a         SNERLHXMPSPQESVXKXFWFMGFSEXSGGLIRERDYCEAVRFDSGALRKRLMLPEKNXP
                      130       140       150       160       170       180

                      190       200       210       220       230       240
orf32-1.pep    EWLLFGYRSDVWAKWLEMWRQAGSPMTLLLAGTQIIDSLKQSGVIPQDALQNDGDVFQTA
               |||||||||||||||||||||||||:||||||: |||||||:||||||||||||||||||
orf32a         EWLLFGYRSDVWAKWLEMWRQAGSPLTLLLAGAXIIDSLKQNGVIPQDALQNDGDVFQTA
                      190       200       210       220       230       240

                      250       260       270       280       290       300
orf32-1.pep    SVRLVKIPFVPQQDFDQLLHLADCAVIRGEDSFVRAQLAGKPFFWHIYPQDENVHLDKLH
               ||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
orf32a         SVRLVKIPFVPQQDFDKLLHLADCAVIRGEDSFVRAQLAGKPFFWHIYPQDENVHLDKLH
                      250       260       270       280       290       300

                      310       320       330       340       350       360
orf32-1.pep    AFWDKAHGFYTPETVSAHRRLSDDLNGGEALSATQRLECWQTLQQHQNGWRQGAEDWSRY
               ||||||||||||||:|||||||||||||||||||||||||| ||||||||||||||||||
orf32a         AFWDKAHGFYTPETASAHRRLSDDLNGGEALSATQRLECWQILQQHQNGWRQGAEDWSRY
                      310       320       330       340       350       360

                      370       380
orf32-1.pep    LFGQPSAPEKLAAFVSKHQKIRX
               ||||||| |||||||||||||||
orf32a         LFGQPSASEKLAAFVSKHQKIRX
                      370       380
```

<u>Homology with a predicted ORF from *N.gonorrhoeae*</u>

ORF32 shows 95.1% identity over a 82aa overlap with a predicted ORF (ORF32.ng) from *N. gonorrhoeae:*

```
orf32.pep    MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLP    57
             |||   | ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
orf32ng      MVMNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLP   60

orf32.pep    DVPCVHQDIHVRTWHSDAADIDTA                                       81
             ||| ||||||||||||||||||||
orf32ng      DVPFVHQDIHVRTWHSDAADIDTAPVPDAVIETFACDLPENVLNIIRRHKPLWLNWEYLS  120
```

An ORF32ng nucleotide sequence <SEQ ID 193> was predicted to encode a protein having amino acid sequence <SEQ ID 194>:

```
  1  MVMNTYAFPV CWIFCKVIDN FGDIGVSWRL ARVLHRELGW QVHLWTDDVS
 51  ALRALCPDLP DVPFVHQDIH VRTWHSDAAD IDTAPVPDAV IETFACDLPE
101  NVLNIIRRHK PLWLNWEYLS AEESNERLHL MPSPQEGVQK YFWFMGFSEK
151  SGGLIRERDY REAVRFDTEA LRRRLVLPEK NAPEWLLFGY RGDVWAKWLD
201  MWQQAGSLMT LLLAGAQIID SLKQSGVIPQ NALQNEGGVF QTASVRLVKI
251  PFVPQQDFDK LLHLADCAVI RGEDSFVRTQ LAGKPFFWHI YPQDENVHLD
301  KLHAFWDKAY GFYTPETASV HRLLSDDLNG GEALSATQRL ECGVL*
```

Further sequencing revealed the following DNA sequence <SEQ ID 195>:

```
   1  ATGAATACAT ACGCTTTTCC TGTCTGTTGG ATTTTTTGCA AGGTCATCGA
  51  CAATTTCGGC GACATCGGCG TTTCGTGGCG GCTCGCCCGT GTTTTGCACC
 101  GCGAACTCGG TTGGCAGGTG CATTTGTGGA CGGACGACGT GTCCGCCTTG
 151  CGCGCGCTTT GTCCCGATTT GCCCGATGTT CCCTTCGTTC ATCAGGATAT
 201  TCATGTCCGC ACTTGGCATT CCGATGCGGC AGACATTGAT ACCGCGCCCG
 251  TTCCCGATGC CGTTATCGAA ACTTTTGCCT GCGACCTGCC CGAAAATGTG
 301  CTGAACATCA TCCGCCGACA CAAACCGCTT TGGCTGAATT GGGAATATTT
 351  GAGCGCGGAG GAAAGCAATG AAAGGCTGCA CCTGATGCCT TCGCCGCAGG
 401  AGGGCGTTCA AAAATATTTT TGGTTTATGG GTTTCAGCGA AAAAAGCGGC
 451  GGGTTGATAC GCGAACGCGA TTACCGCGAA GCCGTCCGTT TCGATACCGA
 501  AGCCCTGCGC CGGCGGCTGG TGCTGCCCGA AAAAAACGCC CCCGAATGGC
 551  TGCTTTTCGG CTATCGGGGC GATGTTTGGG CAAAGTGGCT GGACATGTGG
 601  CAACAGGCAG GCAGCCTGAT GACCCTACTG CTGGCGGGGG CGCAAATTAT
 651  CGACAGCCTC AAACAAAGCG GCGTTATTCC GCAAAACGCC CTGCAAAAtg
 701  aaggcgGTGT CTTTCagacG gcatccgTcC gccttGTCAA AAtcCCGTTC
 751  GTGCcGCAAC AGGAcTTCGA CAAATTGCTG CAcctcgcCG ACTGCGCCGT
 801  GATACGCGGC GAAGACAGTT TCGTGCGTAC CCAGCTTGCC GGAAAACCCT
 851  TTTTTTGGCA CATCTACCCG CAAGACGAGA ATGTCCATCT CGACAAACTC
 901  CACGCCTTTT GGGATAAGGC ATACGGCTTC TACACGCCCG AAACCGCATC
 951  GGTGCACCGC CTCCTTTCGG ACGACCTCAA CGGCGGAGAG GCTTTATCCG
1001  CAACACAACG CCTCGAATGT TGGCAAACCC TGCAACAACA TCAAAACGGC
1051  TGGCGGCAAG GCGCGGAGGA TTGGAGCCGT TATCTTTTCG GGCAGCCTTC
1101  CGCATCCGAA AAACTCGCCG CCTTTGTTTC AAAGCATCAA AAAATACGCT
1151  AG
```

This encodes a protein having amino acid sequence <SEQ ID 196; ORF32ng-1>:

```
  1  MNTYAFPVCW IFCKVIDNFG DIGVSWRLAR VLHRELGWQV HLWTDDVSAL
 51  RALCPDLPDV PFVHQDIHVR TWHSDAADID TAPVPDAVIE TFACDLPENV
101  LNIIRRHKPL WLNWEYLSAE ESNERLHLMP SPQEGVQKYF WFMGFSEKSG
151  GLIRERDYRE AVRFDTEALR RRLVLPEKNA PEWLLFGYRG DVWAKWLDMW
201  QQAGSLMTLL LAGAQIIDSL KQSGVIPQNA LQNEGGVFQT ASVRLVKIPF
251  VPQQDFDKLL HLADCAVIRG EDSFVRTQLA GKPFFWHIYP QDENVHLDKL
301  HAFWDKAYGF YTPETASVHR LLSDDLNGGE ALSATQRLEC WQTLQQHQNG
351  WRQGAEDWSR YLFGQPSASE KLAAFVSKHQ KIR*
```

ORF32ng-1 and ORF32-1 show 93.5% identity in 383 aa overlap:

```
                         10        20        30        40        50         59
orf32-1.pep     MNTPPF-VCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDV
                |||  | |||||||||||||||||||||||||||||||||||||||||||||||||||||
orf32ng-1       MNTYAFPVCWIFCKVIDNFGDIGVSWRLARVLHRELGWQVHLWTDDVSALRALCPDLPDV
                        10        20        30        40        50        60

               60        70        80        90       100       110       119
orf32-1.pep     PCVHQDIHVRTWHSDAADIDTAPVPDVVIETFACDLPENVLHIIRRHKPLWLNWEYLSAE
                | ||||||||||||||||||||||||:|||||||||||||:|||||||||||||||||||
orf32ng-1       PFVHQDIHVRTWHSDAADIDTAPVPDAVIETFACDLPENVLNIIRRHKPLWLNWEYLSAE
                        70        80        90       100       110       120

              120       130       140       150       160       170       179
```

```
orf32-1.pep    ESNERLHLMPSPQEGVQKYFWFMGFSEKSGGLIRERDYCEAVRFDTEALRERLMLPEKNA
               |||||||||||||||||||||||||||||||||||||| |||||||||||:||:||||||
orf32ng-1      ESNERLHLMPSPQEGVQKYFWFMGFSEKSGGLIRERDYREAVRFDTEALRRRLVLPEKNA
                       130       140       150       160       170       180

              180       190       200       210       220       230       239
orf32-1.pep    SEWLLFGYRSDVWAKWLEMWRQAGSPMTLLLAGTQIIDSLKQSGVIPQDALQNDGDVFQT
                ||||||||:|||||||:||:|||| |||||||:||||||||||||||:||||:| ||||
orf32ng-1      PEWLLFGYRGDVWAKWLDMWQQAGSLMTLLLAGAQIIDSLKQSGVIPQNALQNEGGVFQT
                       190       200       210       220       230       240

              240       250       260       270       280       290       299
orf32-1.pep    ASVRLVKIPFVPQQDFDQLLHLADCAVIRGEDSFVRAQLAGKPFFWHIYPQDENVHLDKL
               |||||||||||||||||:||||||||||||||||||:|||||||||||||||||||||||
orf32ng-1      ASVRLVKIPFVPQQDFDKLLHLADCAVIRGEDSFVRTQLAGKPFFWHIYPQDENVHLDKL
                       250       260       270       280       290       300

              300       310       320       330       340       350       359
orf32-1.pep    HAFWDKAHGFYTPETVSAHRRLSDDLNGGEALSATQRLECWQTLQQHQNGWRQGAEDWSR
               |||||||:|||||||:|:|| |||||||||||||||||||||||||||||||||||||||
orf32ng-1      HAFWDKAYGFYTPETASVHRLLSDDLNGGEALSATQRLECWQTLQQHQNGWRQGAEDWSR
                       310       320       330       340       350       360

              360       370       380
orf32-1.pep    YLFGQPSAPEKLAAFVSKHQKIRX
               |||||||| |||||||||||||||
orf32ng-1      YLFGQPSASEKLAAFVSKHQKIRX
                       370       380
```

On this basis, including the RGD sequence in the gonococcal protein, characteristic of adhesins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF32-1 (42kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figure 7A shows the results of affinity purification of the His-fusion protein, and Figure 7B shows the results of expression of the GST-fusion in *E.coli*. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA, giving a positive result. These experiments confirm that ORF32-1 is a surface-exposed protein, and that it is a useful immunogen.

Example 24

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 197>:

```
  1  ..TTGTTCCTGC GTGTNAAAGT GGGGCGTTTT TTCAGCAGTC CGGCGACGTG
 51    GTTTCGGGNC AAAGACCCTG TAAATCAGGC GGTGTTGCGG CTGTATNCGG
101    ACGAGTGGCG GCA.ACTTCG GTACGTTGGA AAATAGNCGC AACGTCGCAC
151    AGCCTGTGGC TCTGCACGCT GCTCGGAATG CTGGTGTCGG TATTGTTGCT
201    GCTTTTGGTG CGGCAATATA CGTTCAACTG GGAAAGCACG CTGTTGAGCA
251    ATGCCGCTTC GGTACGCGCG GTGGAAATGT TGGCATGGCT GCCGTCGAAA
301    CTCGGTTTCC CTGTCCCCGA TGCGCGGTCG GTCATCGAAG GCCGTCTGAA
351    CGGCAATATT GCCGATGCGC GGGCTTGGTC GGGGCTGCTG GTCGNCAGTA
401    TCGCCTGCTA NGGCATCCTG CCGCGCCTG..
```

This corresponds to the amino acid sequence <SEQ ID 198; ORF33>:

```
  1  ..LFLRVKVGRF FSSPATWFRX KDPVNQAVLR LYXDEWRXTS VRWKIXATSH
 51    SLWLCTLLGM LVSVLLLLLV RQYTFNWEST LLSNAASVRA VEMLAWLPSK
101    LGFPVPDARS VIEGRLNGNI ADARAWSGLL VXSIACXGIL PRL..
```

Further work revealed the complete nucleotide sequence <SEQ ID 199>:

```
   1  ATGTTGAATC CATCCCGAAA ACTGGTTGAG CTGGTCCGTA TTTTGGACGA
  51  AGGCGGTTTT ATTTTCAGCG GCGATCCCGT ACAGGCGACG GAGGCTTTGC
 101  GCCGCGTGGA CGGCAGTACG GAGGAAAAAA TCATCCGTCG GGCGGAGATG
 151  ATTGACAGGA ACCGTATGCT GCGGGAGACG TTGGAACGTG TGCGTGCGGG
 201  GTCGTTCTGG TTGTGGGTGG TGGCGGCGAC GTTTGCATTT TTTACCGGTT
 251  TTTCAGTCAC TTATCTTCTA ATGGACAATC AGGGTCTGAA TTTCTTTTTG
 301  GTTTTGGCGG GCGTGTTGGG CATGAATACG CTGATGCTGG CAGTATGGTT
 351  GGCAATGTTG TTCCTGCGTG TGAAAGTGGG GCGTTTTTTC AGCAGTCCGG
 401  CGACGTGGTT TCGGGGCAAA GACCCTGTAA ATCAGGCGGT GTTGCGGCTG
 451  TATGCGGACG AGTGGCGGCA ACCTTCGGTA CGTTGGAAAA TAGGCGCAAC
 501  GTCGCACAGC CTGTGGCTCT GCACGCTGCT CGGAATGCTG GTGTCGGTAT
 551  TGTTGCTGCT TTTGGTGCGG CAATATACGT TCAACTGGGA AAGCACGCTG
 601  TTGAGCAATG CCGCTTCGGT ACGCGCGGTG GAAATGTTGG CATGGCTGCC
 651  GTCGAAACTC GGTTTCCCTG TCCCCGATGC GCGGGCGGTC ATCGAAGGCC
 701  GTCTGAACGG CAATATTGCC GATGCGCGGG CTTGGTCGGG GCTGCTGGTC
 751  GGCAGTATCG CCTGCTACGG CATCCTGCCG CGCCTGCTGG CTTGGGTAGT
 801  GTGTAAAATC CTTTTGAAAA CAAGCGAAAA CGGATTGGAT TTGGAAAAGC
 851  CCTATTATCA GGCGGTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG
 901  GATACGCGTC GGGAAACCGT GTCCGCCGTT TCACCGAAAA TCATCTTGAA
 951  CGATGCGCCG AAATGGGCGG TCATGCTGGA GACCGAGTGG CAGGACGGCG
1001  AATGGTTCGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
1051  ACCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
1101  GGCGCAACTG CTTATCGGCG TGCGCGCCCA AACTGTGCCG GACCGCGGCG
1151  TGTTGCGGCA GATTGTCCGA CTCTCGGAAG CGGCGCAGGG CGGCGCGGTG
1201  GTGCAGCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT
1251  GGAACATTGG CGTAACGCGC TGGCCGAATG CGGCGCGGCG TGGCTTGAGC
1301  CTGACAGGGC GGCGCAGGAA GGGCGTTTGA AAGACCAATA A
```

This corresponds to the amino acid sequence <SEQ ID 200; ORF33-1>:

```
  1  MLNPSRKLVE LVRILDEGGF IFSGDPVQAT EALRRVDGST EEKIIRRAEM
 51  IDRNRMLRET LERVRAGSFW LWVVAATFAF FTGFSVTYLL MDNQGLNFFL
101  VLAGVLGMNT LMLAVWLAML FLRVKVGRFF SSPATWFRGK DPVNQAVLRL
151  YADEWRQPSV RWKIGATSHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL
201  LSNAASVRAV EMLAWLPSKL GFPVPDARAV IEGRLNGNIA DARAWSGLLV
251  GSIACYGILP RLLAWVVCKI LLKTSENGLD LEKPYYQAVI RRWQNKITDA
301  DTRRETVSAV SPKIILNDAP KWAVMLETEW QDGEWFEGRL AQEWLDKGVA
351  TNREQVAALE TELKQKPAQL LIGVRAQTVP DRGVLRQIVR LSEAAQGGAV
401  VQLLAEQGLS DDLSEKLEHW RNALAECGAA WLEPDRAAQE GRLKDQ*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF33 shows 90.9% identity over a 143aa overlap with an ORF (ORF33a) from strain A of *N. meningitidis:*

```
                                                   10        20        30
orf33.pep                                 LFLRVKVGRFFSSPATWFRXKDPVNQAVLR
                                          ||||||||||||||||||| ||||||||||
orf33a        LMDNQGLNFFLVLAGVXGMNTLMLAVWLAMLFLRVKVGRFFSSPATWFRGKDPVNQAVLR
              90       100       110       120       130       140

                      40        50        60        70        80        90
orf33.pep     LYXDEWRXTSVRWKIXATSHSLWLCTLLGMLVSVLLLLLVRQYTFNWESTLLSNAASVRA
              || ||||| |||||| ||||||||||||||||||||||||||||||||||||::::|||
orf33a        LYADEWRXPSVRWKIGATSHSLWLCTLLGMLVSVLLLLLVRQYTFNWESTLLGDSSSVRL
             150       160       170       180       190       200

                     100       110       120       130       140
orf33.pep     VEMLAWLPSKLGFPVPDARSVIEGRLNGNIADARAWSGLLVXSIACXGILPRL
              ||||||||:||||||||||:||||||||||||||||||||| |||| ||||||
orf33a        VEMLAWLPAKLGFPVPDARAVIEGRLNGNIADARAWSGLLVGSIACYGILPRLLAWAVCK
             210       220       230       240       250       260

orf33a        ILXXTSENGLDLEKXXXXXXIRRWQNKITDADTRRETVSAVSPKIVLNDAPKWAVMLETE
             270       280       290       300       310       320
```

The complete length ORF33a nucleotide sequence <SEQ ID 201> is:

```
   1  ATGTTGAATC CATCCCGAAA ACTGGTTGAG CTGGTCCGTA TTTTGGAAGA
  51  AGGCGGCTTT ATTTTCAGCG GCGATCCCGT GCAGGCGACG GAGGCTTTGC
 101  GCCGCGTGGA CGGCAGTACG GAGGAAAAAA TCATCCGTCG GGCGAAGATG
 151  ATCGACAGGA ACCGTATGCT GCGGGAGACG TTGGAACGTG TGCGTGCGGG
 201  GTCGTTCTGG TTGTGGGTGG CGGCGGCGAC GTTTGCGTTT NTTACCGNTT
 251  TTTCAGTTAC TTATCTTCTA ATGGACAATC AGGGTCTGAA TTTCTTTTTG
 301  GTTTTGGCGG GCGTGNTGGG CATGAATACG CTGATGCTGG CAGTATGGTT
 351  GGCAATGTTG TTCCTGCGCG TGAAAGTGGG GCGTTTTTTC AGCAGTCCGG
 401  CGACGTGGTT TCGGGGCAAA GACCCTGTCA ATCAGGCGGT GTTGCGGCTG
 451  TATGCGGACG AGTGGCGGCN ACCTTCGGTA CGTTGGAAAA TAGGCGCAAC
 501  GTCGCACAGC CTGTGGCTCT GCACGCTGCT CGGAATGCTG GTGTCGGTAT
 551  TGTTGCTGCT TTTGGTGCGG CAATATACGT TCAACTGGGA AAGCACGCTG
 601  TTGGGCGATT CGTCTTCGGT ACGGCTGGTG GAAATGTTGG CATGGCTGCC
 651  TGCGAAACTG GGTTTTCCCG TGCCTGATGC GCGGGCGGTC ATCGAAGGTC
 701  GTCTGAACGG CAATATTGCC GATGCGCGGG CTTGGTCGGG GCTGCTGGTC
 751  GGCAGTATCG CCTGCTACGG CATCCTGCCG CGCCTCTTGG CTTGGGCGGT
 801  ATGCAAAATC CTTNTGNAAA CAAGCGAAAA CGGCTTGGAT TTGGAAAAGC
 851  NCNNNNNTCN NNCGNTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG
 901  GATACGCGTC GGGAAACCGT GTCCGCCGTT TCGCCGAAAA TCGTCTTGAA
 951  CGATGCGCCG AAATGGGCGG TCATGCTGGA GACCGAATGG CAGGACGGCG
1001  AATGGTTCGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
1051  GCCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
1101  GGCGCAACTG CTTATCGGCG TGCGCGCCCA AACTGTGCCC GACCGCGGCG
1151  TGTTGCGGCA GATCGTCCGA CTTTCGGAAG CGGCGCAGGG CGGCGCGGTG
1201  GTGCANCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT
1251  GGAACATTGG CGTAACGCGC TGACCGAATG CGGCGCGGCG TGGCTGGAAC
1301  CCGACAGAGC GGCGCAGGAA GGCCGTCTGA AAACCAACGA CCGCACTTGA
```

This encodes a protein having amino acid sequence <SEQ ID 202>:

```
  1  MLNPSRKLVE LVRILEEGGF IFSGDPVQAT EALRRVDGST EEKIIRRAKM
 51  IDRNRMLRET LERVRAGSFW LWVAAATFAF XTXFSVTYLL MDNQGLNFFL
101  VLAGVXGMNT LMLAVWLAML FLRVKVGRFF SSPATWFRGK DPVNQAVLRL
151  YADEWRXPSV RWKIGATSHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL
201  LGDSSSVRLV EMLAWLPAKL GFPVPDARAV IEGRLNGNIA DARAWSGLLV
251  GSIACYGILP RLLAWAVCKI LXXTSENGLD LEKXXXXXXI RRWQNKITDA
301  DTRRETVSAV SPKIVLNDAP KWAVMLETEW QDGEWFEGRL AQEWLDKGVA
351  ANREQVAALE TELKQKPAQL LIGVRAQTVP DRGVLRQIVR LSEAAQGGAV
401  VXLLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRAAQE GRLKTNDRT*
```

ORF33a and ORF33-1 show 94.1% identity in 444 aa overlap:

```
                       10        20        30        40        50        60
orf33a.pep    MLNPSRKLVELVRILEEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAKMIDRNRMLRET
              ||||||||||||||:||||||||||||||||||||||||||||||||||:||||||||||
orf33-1       MLNPSRKLVELVRILDEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAEMIDRNRMLRET
                       10        20        30        40        50        60

                       70        80        90       100       110       120
orf33a.pep    LERVRAGSFWLWVAAATFAFXTXFSVTYLLMDNQGLNFFLVLAGVXGMNTLMLAVWLAML
              |||||||||||||:|||||| | |||||||||||||||||||||| ||||||||||||||
orf33-1       LERVRAGSFWLWVVAATFAFFTGFSVTYLLMDNQGLNFFLVLAGVLGMNTLMLAVWLAML
                       70        80        90       100       110       120

                      130       140       150       160       170       180
orf33a.pep    FLRVKVGRFFSSPATWFRGKDPVNQAVLRLYADEWRXPSVRWKIGATSHSLWLCTLLGML
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
orf33-1       FLRVKVGRFFSSPATWFRGKDPVNQAVLRLYADEWRQPSVRWKIGATSHSLWLCTLLGML
                      130       140       150       160       170       180

                      190       200       210       220       230       240
orf33a.pep    VSVLLLLLVRQYTFNWESTLLGDSSSVRLVEMLAWLPAKLGFPVPDARAVIEGRLNGNIA
              |||||||||||||||||||||::::||| ||||||||:||||||||||||||||||||||
orf33-1       VSVLLLLLVRQYTFNWESTLLSNAASVRAVEMLAWLPSKLGFPVPDARAVIEGRLNGNIA
                      190       200       210       220       230       240

                      250       260       270       280       290       300
orf33a.pep    DARAWSGLLVGSIACYGILPRLLAWAVCKILXXTSENGLDLEKXXXXXXIRRWQNKITDA
              |||||||||||||||||||||||||:|||||  ||||||||||      |||||||||||
```

```
orf33-1     DARAWSGLLVGSIACYGILPRLLAWVVCKILLKTSENGLDLEKPYYQAVIRRWQNKITDA
                    250       260       270       280       290       300

                    310       320       330       340       350       360
orf33a.pep  DTRRETVSAVSPKIVLNDAPKWAVMLETEWQDGEWFEGRLAQEWLDKGVAANREQVAALE
            ||||||||||||||:|||||||||||||||||||||||||||||||||||:|||||||||
orf33-1     DTRRETVSAVSPKIILNDAPKWAVMLETEWQDGEWFEGRLAQEWLDKGVATNREQVAALE
                    310       320       330       340       350       360

                    370       380       390       400       410       420
orf33a.pep  TELKQKPAQLLIGVRAQTVPDRGVLRQIVRLSEAAQGGAVVXLLAEQGLSDDLSEKLEHW
            ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
orf33-1     TELKQKPAQLLIGVRAQTVPDRGVLRQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW
                    370       380       390       400       410       420

                    430       440       450
orf33a.pep  RNALTECGAAWLEPDRAAQEGRLKTNDRTX
            ||||:|||||||||||||||||||
orf33-1     RNALAECGAAWLEPDRAAQEGRLKDQX
                    430       440
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF33 shows 91.6% identity over a 143aa overlap with a predicted ORF (ORF33.ng) from *N. gonorrhoeae:*

```
orf33.pep                                 LFLRVKVGRFFSSPATWFRXKDPVNQAVLR   30
                                          ||||||||||||||||||| | ||||||||
orf33ng     LMDNQGLNFFLVLAGVLGMNTLMLAVWLATLFLRVKVGRFFSSPATWFRGKGPVNQAVLR  100

orf33.pep   LYXDEWRXTSVRWKIXATSHSLWLCTLLGMLVSVLLLLLVRQYTFNWESTLLSNAASVRA   90
            || |:||  |||||| ||:|||||||||||||||||||||||||||||||||||||||||
orf33ng     LYADQWRQPSVRWKIGATAHSLWLCTLLGMLVSVLLLLLVRQYTFNWESTLLSNAASVRA  160

orf33.pep   VEMLAWLPSKLGFPVPDARSVIEGRLNGNIADARAWSGLLVXSIACXGILPRL         143
            |||||||||||||||||||:||||||||||||||||||||| ||:| ||||||
orf33ng     VEMLAWLPSKLGFPVPDARAVIEGRLNGNIADARAWSGLLVGSIVCYGILPRLLAWVVCK  220
```

An ORF33ng nucleotide sequence <SEQ ID 203> was predicted to encode a protein having amino acid sequence <SEQ ID 204>:

```
  1  MIDRDRMLRD TLERVRAGSF WLWVVVASMM FTAGFSGTYL LMDNQGLNFF
 51  LVLAGVLGMN TLMLAVWLAT LFLRVKVGRF FSSPATWFRG KGPVNQAVLR
101  LYADQWRQPS VRWKIGATAH SLWLCTLLGM LVSVLLLLLV RQYTFNWEST
151  LLSNAASVRA VEMLAWLPSK LGFPVPDARA VIEGRLNGNI ADARAWSGLL
201  VGSIVCYGIL PRLLAWVVCK ILLKTSENGL DLEKTYYQAV IRRWQNKITD
251  ADTRRETVSA VSPKIVLNDA PKWALMLETE WQDGQWFEGR LAQEWLDKGV
301  AANREQVAAL ETELKQKPAQ LLIGVRAQTV PDRGVLRQIV RLSEAAQGGA
351  VVQLLAEQGL SDDLSEKLEH WRNALTECGA AWLEPDRVAQ EGRLKDQ*
```

Further sequence analysis revealed the following DNA sequence <SEQ ID 205>:

```
  1  ATGTTGaatC CATCCCgaAA ACTGgttgag ctGgTCCgtA Ttttgaataa
 51  agggggtTTT attttcagcg gcgatcctgt gcaggcgacg gaggctttgc
101  gccgcgtgga cggcAGTACG GAggAaaaaa tcttccgtcg GGCGGAGAtg
151  atcgACAGGg accgtatgtt gcgggACaCg TtggaacGTG TGCGTGCggg
201  gtcgtTctgG TTATGGGTGG TggtggCAtC gATGATGTtt aCCGCCGGAT
251  TTTCAGgcac ttatCttCTG ATGGACaatC AGGGGCtGAA TtTCTTTTTA
301  GTTTTggcgG GAGTGTtggG CATGaatacG ctgATGCTGG CAGTATGGtt
351  gGCAACGTTG TTCCTGCGCG TGAAAGTGGG ACGGTTTTTC AGCAGTCCGG
401  CGACGTGGTT TCGGGGCAAA GGCCCTGTAA ATCAGGCGGT GTTGCGGCTG
451  TATGCGGACC AGTGGCGGCA ACCTTCGGTA CGATGGAAAA TAGGCGCAAC
501  GGCGCACAGC TTGTGGCTCT GCACGCTGCT CGGAATGCTG GTGTCGGTAT
551  TGCTGCTGCT TTTGGTGCGG CAATATACGT TCAACTGGGA AAGCACGCTG
601  TTGAGCAATG CCGCTTCGGT ACGCGCGGTG GAAATGTTGG CATGGCTGCC
651  GTCGAAACTC GGTTTCCCTG TCCCCGATGC GCGGGCGGTC ATCGAAGGTC
701  GTCTGAACGG CAATATTGCC GATGCGCGGG CTTGGTCGGG GCTGCTGGTC
751  GGCAGTATCG TCTGCTACGG CATCCTGCCG CGCCTCTTGG CTTGGGTAGT
```

```
 801  GTGTAAAATC CTTTTGAAAA CAAGCGAAAA CGGattgGAT TTGGAAAAAA
 851  CCTATTATCA GGCGGTCATC CGCCGCTGGC AGAACAAAAT CACCGATGCG
 901  GATACGCGTC GGGAAACCGT GTCCGCCGTT TCGCcgaAAA TCGTCTTGAA
 951  CGATGCGCCG AAATGGGCGC TCATGCTGGA GACCGAGTGG CAGGACGGCC
1001  AATGGTTCGA GGGCAGGCTG GCGCAGGAAT GGCTGGATAA GGGCGTTGCC
1051  GCCAATCGGG AACAGGTTGC CGCGCTGGAG ACAGAGCTGA AGCAGAAACC
1101  GGCGCAACTG CTTATCGGCG TACGCGCCCA AACTGTGCCG GACCGGGGCG
1151  TGCTGCGGCA GATTGTGCGG CTTTCGGAAG CGGCGCAGGG CGGCGCGGTG
1201  GTGCAGCTTT TGGCGGAACA GGGGCTTTCA GACGACCTTT CGGAAAAGCT
1251  GGAACATTGG CGTAACGCGC TGACCGAATG CGGCGCGGCG TGGCTTGAGC
1301  CTGACAGGGT GGCGCAGGAA GGCCGTTTGA AAGACCAATA A
```

This encodes a protein having amino acid sequence <SEQ ID 206; ORF33ng-1>:

```
  1  MLNPSRKLVE LVRILNKGGF IFSGDPVQAT EALRRVDGST EEKIFRRAEM
 51  IDRDRMLRDT LERVRAGSFW LWVVVASMMF TAGFSGTYLL MDNQGLNFFL
101  VLAGVLGMNT LMLAVWLATL FLRVKVGRFF SSPATWFRGK GPVNQAVLRL
151  YADQWRQPSV RWKIGATAHS LWLCTLLGML VSVLLLLLVR QYTFNWESTL
201  LSNAASVRAV EMLAWLPSKL GFPVPDARAV IEGRLNGNIA DARAWSGLLV
251  GSIVCYGILP RLLAWVVCKI LLKTSENGLD LEKTYYQAVI RRWQNKITDA
301  DTRRETVSAV SPKIVLNDAP KWALMLETEW QDGQWFEGRL AQEWLDKGVA
351  ANREQVAALE TELKQKPAQL LIGVRAQTVP DRGVLRQIVR LSEAAQGGAV
401  VQLLAEQGLS DDLSEKLEHW RNALTECGAA WLEPDRVAQE GRLKDQ*
```

ORF33ng-1 and ORF33-1 show 94.6% identity in 446 aa overlap:

```
                       10        20        30        40        50        60
orf33-1.pep    MLNPSRKLVELVRILDEGGFIFSGDPVQATEALRRVDGSTEEKIIRRAEMIDRNRMLRET
               |||||||||||||||::|||||||||||||||||||||||||||:||||||||:||||:|
orf33ng-1      MLNPSRKLVELVRILNKGGFIFSGDPVQATEALRRVDGSTEEKIFRRAEMIDRDRMLRDT
                       10        20        30        40        50        60

                       70        80        90       100       110       120
orf33-1.pep    LERVRAGSFWLWVVAATFAFFTGFSVTYLLMDNQGLNFFLVLAGVLGMNTLMLAVWLAML
               ||||||||||||||:|:: | :||| |||||||||||||||||||||||||||||||| |
orf33ng-1      LERVRAGSFWLWVVVASMMFTAGFSGTYLLMDNQGLNFFLVLAGVLGMNTLMLAVWLATL
                       70        80        90       100       110       120

                      130       140       150       160       170       180
orf33-1.pep    FLRVKVGRFFSSPATWFRGKDPVNQAVLRLYADEWRQPSVRWKIGATSHSLWLCTLLGML
               |||||||||||||||||||| |||||||||||| ||||||||||||| ||||||||||||
orf33ng-1      FLRVKVGRFFSSPATWFRGKGPVNQAVLRLYADQWRQPSVRWKIGATAHSLWLCTLLGML
                      130       140       150       160       170       180

                      190       200       210       220       230       240
orf33-1.pep    VSVLLLLLVRQYTFNWESTLLSNAASVRAVEMLAWLPSKLGFPVPDARAVIEGRLNGNIA
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf33ng-1      VSVLLLLLVRQYTFNWESTLLSNAASVRAVEMLAWLPSKLGFPVPDARAVIEGRLNGNIA
                      190       200       210       220       230       240

                      250       260       270       280       290       300
orf33-1.pep    DARAWSGLLVGSIACYGILPRLLAWVVCKILLKTSENGLDLEKPYYQAVIRRWQNKITDA
               |||||||||||||:||||||||||||||||||||||||||||| ||||||||||||||||
orf33ng-1      DARAWSGLLVGSIVCYGILPRLLAWVVCKILLKTSENGLDLEKTYYQAVIRRWQNKITDA
                      250       260       270       280       290       300

                      310       320       330       340       350       360
orf33-1.pep    DTRRETVSAVSPKIILNDAPKWAVMLETEWQDGEWFEGRLAQEWLDKGVATNREQVAALE
               ||||||||||||||:||||||||:|||||||||:||||||||||||||||:|||||||||
orf33ng-1      DTRRETVSAVSPKIVLNDAPKWALMLETEWQDGQWFEGRLAQEWLDKGVAANREQVAALE
                      310       320       330       340       350       360

                      370       380       390       400       410       420
orf33-1.pep    TELKQKPAQLLIGVRAQTVPDRGVLRQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf33ng-1      TELKQKPAQLLIGVRAQTVPDRGVLRQIVRLSEAAQGGAVVQLLAEQGLSDDLSEKLEHW
                      370       380       390       400       410       420

                      430       440
orf33-1.pep    RNALAECGAAWLEPDRAAQEGRLKDQX
               ||||:|||||||||||:|||||||||
```

```
orf33ng-1       RNALTECGAAWLEPDRVAQEGRLKDQX
                          430       440
```

Based on the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 25

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 207>:

```
  1  ..CAGAAGAGTT TGTCGAGAAT TTCTTTATGG GGTTTGGGCG GCGTGTTTTT
 51    CGGGGTGTCC GGTCTGGTAT GGTTTTCTTT GGGCGTTTCT TT.GAGTGCG
101    CCTGTTTTTC GGGTGTTTCT TTTCGGGGTT CGGGACGGGG GACGTTTGTG
151    GGCAGTACGG GGGTTTCTTT GAGTGTGTTT TCAGCTTGTG TTCC.GGCGT
201    CGTCCGGCTG CCTGTCGGTT TGAGCTGTGT CGGCAGGTTG CG..GTTTGA
251    CCCGGTTTTT CTTGGGTGCG GCAGGGGACG TCATTCTCCT GCCGCTTTCG
301    TCTGTGCCGT CCGGCTGTGC GGGTTCGGAT GAGGCGGCGT GGTGGTGTTC
351    GGGTTGGGCG GCATCTTGTT CCGACTACGC CGTTTGGCAG CCAGAATTCG
401    GTTTCGCGGG GGCTGTCGGT GTGTTGCGGT TCGGCTTGAA GGGTTTTGTC
451    GTCC..
```

This corresponds to the amino acid sequence <SEQ ID 208; ORF34>:

```
  1  ..QKSLSRISLW GLGGVFFGVS GLVWFSLGVS XECACFSGVS FRGSGRGTFV
 51    GSTGVSLSVF SACVXGGVVRL PVGLSCVGRL XXLTRFFLGA AGDVILLPLS
101    SVPSGCAGSD EAAWWCSGWA ASCPTTPFGS QNSVSRGLSV CCGSA*RVLS
151    S..
```

Further work revealed the complete nucleotide sequence <SEQ ID 209>:

```
   1  ATGATGATGC CGTTCATAAT GCTTCCTTGG ATTGCkGGTG TGCCTGCCGT
  51  GCCGGGTCAG AATAGGTTGT CCAGAATTTC TTTATGGGGT TTGGGCGGCG
 101  TGTTTTTCGG GGTGTCCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTTTG
 151  GGCTGCGCCT GTTTTTCGGG TGTTTCTTTT CGGGGTTCGG GACGGGGGAC
 201  GTTTGTGGGC AGTACGGGGG TTTCTTTGAG TGTGTTTTCA GCTTGTGTTC
 251  CGGCGTCGTC CGGCTGCCTG TCGGTTTGAG CTGTGTCGGC AGGTTGCGGT
 301  TTGACCCGGT TTTTCTTGGG TGCGGCAGGG GACGGCAGTC CGCTGCCGCT
 351  TTCGTCTGTG CCGTCCGGCT GTGCGGGTTC GGATGAGGCG GCGTGGTGGT
 401  GTTCGGGTTG GGCGGCATCT TGTCCGACTA CGCCGTTTGG CAGCCAGAAT
 451  TCGGTTTCGC GGGGGCTGTC GGTGTGTTGC GGTTCGGCTT GAAGGGTTTT
 501  GTCGCCGTTC GGGTTGAATG TGCTGACGAT GCCTATTGCC AATGCGCCGA
 551  TGGCGGCGAT ACAGATGAGC AATACGGCGC GTATCAGGAG TTTGGGGGTC
 601  AGCCTGAAGG GTTTGTTCGG TTTTTTTGCC ATTTTGATTG TGCTTTTGGG
 651  GTGTCGGGCA ATGCCGTCTG AAGGCGGTTC AGACGGCATT GCCGAGTCAG
 701  CGTTGGACGT AGTTTTGGTA GAGGGTGATG ACTTTTTGTA CGCCGACGGT
 751  GGTGCTGACT TTTTGGCTAA TCTGCGCCTG TTCTTCGGGG GTGAGGATGC
 801  CCATAACGTA GGTTACGTTG CCGTAGGTAA CGATTTTGAC GCGCGCCTGT
 851  GTGGCGGGGC TGATGCCCAA CAGCGTGGCG CGGACTTTGG ATGTGTTCCA
 901  AGTGTCGCCG GCGATGTCGC CGGCAGTGCG CGGCAGGGAG GCGACGGTAA
 951  TATAGTTGTA CACGCCTTCG GCGGCCTGTT CGGAACGTGC AATCTGACCG
1001  ACGAACTGTT TTTCGCCTTC GGTGGCGACT TGTCCGAGCA GCAGCAGGTG
1051  GCGGTTGTAG CCGACGACGG AGATTTGGGG CGTGTAGCCT TTGGTTTGGT
1101  TGTTTTGGCG CAGATAGGAA CGGGCGGTGG TTTCGATACG CAACGCCATA
1151  ACGTTGTCGT CGGTTTGCGC GCCGGTGGTT CGGCGGTCGA CGGCGGATTT
1201  CGCGCCCGACG GCGGCGCTTC CGATTACTGC GCTGACGCAG CCGCTAAGGG
1251  CAAGGCTGAA AATGGCGGCA ATCAGGGTGC GGACGGTGTG CGGTTTGGGT
1301  TTCATCGGGT GCTTCCTTTC TTGGGCGTTT CAGACGGCAT TGCTTTGCGC
1351  CATGCCGTCT GA
```

This corresponds to the amino acid sequence <SEQ ID 210; ORF34-1>:

```
  1  MMMPFIMLPW IAGVPAVPGQ NRLSRISLWG LGGVFFGVSG LVWFSLGVSL
 51  GCACFSGVSF RGSGRGTFVG STGVSLSVFS ACVPASSGCL SV*AVSAGCG
101  LTRFFLGAAG DGSPLPLSSV PSGCAGSDEA AWWCSGWAAS CPTTPFGSQN
151  SVSRGLSVCC GSA*RVLSPF GLNVLTMPIA NAPMAAIQMS NTARIRSLGV
```

```
201  SLKGLFGFFA ILIVLLGCRA MPSEGGSDGI AESALDVVLV EGDDFLYADG
251  GADFLGNLRL FFGGEDAHNV GYVAVGNDFD ARLCGGADAQ QRGADFGCVP
301  SVAGDVAGSA RQGGDGNIVV HAFGGLFGTC NLTDELFFAF GGDLSEQQQV
351  AVVADDGDLG RVAFGLVVLA QIGTGGGFDT QRHNVVVGLR AGGSAVDGGF
401  RADGGASDYC ADAAAKGKAE NGGNQGADGV RFGFHRVLPF LGVSDGIALR
451  HAV*
```

Computer analysis of this amino acid sequence gave the following results:

<u>Homology with a predicted ORF from *N.meningitidis* (strain A)</u>

ORF34 shows 73.3% identity over a 161aa overlap with an ORF (ORF34a) from strain A of *N. meningitidis:*

```
                                          10        20        30
orf34.pep                         QKSLSRISLWGLGGVFFGVSGLVWFSLGVSXE------CAC
                                  || ||| ||||||| |||||||||||||||        |||
orf34a         MMXPXIMLPWIAGVPAVPGQKRLSRXSLWGLGGXFFGVSGLVWFSLGVSXSLGVSXGCAC
                       10        20        30        40        50        60

                 40        50        60        70        80              90
orf34.pep      FSGVSFRGSGRGTFVGSTGVSLSVFSACVXGVVRLPVGLSCVGRLXX-----LTRFFLGA
               ||||||||||||||||||||||||||||:      |:: :|::         ||| | ||
orf34a         FSGVSFRGSGRGTFVGSTGVSLSVFSACA------PASSGCLSVXAVSAGCGLTRXFXGA
                       70        80              90       100       110

                     100       110       120       130       140        150
orf34.pep      AGDVILLPLSSVPSGCAGSDEAAWWCSGWAASCPTTPFGSQNSVSRGLSVCCGSAXRVLS
               |||   ||||||||||||:|| |  |||||||||||||||||||||||||||||: ||||
orf34a         AGDGSPLPLSSVPSGCAGADEEAXXCSGWAASCPTTPFGSQNSVSRGLSVCCGSVWRVLS
                 120       130       140       150       160       170

orf34.pep      S

orf34a         PFGXNVLTMPIANAPMAVIQMSNTARIRSLGVSLKGLFXFFAILIVLLGCRAMPSEGGSD
                 180       190       200       210       220       230
```

The complete length ORF34a nucleotide sequence <SEQ ID 211> is:

```
   1  ATGATGATNC CGTTNATAAT GCTTCCTTGG ATTGCGGGTG TGCCTGCCGT
  51  GCCGGGTCAG AAGAGGTTGT CGAGAANTTC TTTATGGGGT TTAGGCGGCN
 101  TGTTTTTCGG GGTGTCCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTNTT
 151  TCTTTGGGTG TTTCTNTGGG CTGTGCCTGT TTTTCGGGTG TTTCTTTTCG
 201  GGGTTCGGGA CGGGGGACGT TTGTGGGCAG TACNGGGGTT TCTTTGAGTG
 251  TGTTTTCAGC TTGTGCTCCG GCGTCGTCCG GCTGCCTGTC GGTTTNAGCT
 301  GTGTCGGCAG GTTGCGGTTT GACCCGGNTT TTCTTNGGTG CGGCAGGGGA
 351  CGGCAGTCCG CTGCCGCTTT CGTCTGTGCC GTCCGGCTGT GCGGGTGCGG
 401  ATGAGGAGGC GTNGTNGTGT TCGGGTTGGG CGGCATCTTG TCCGACTACG
 451  CCGTTTGGCA GCCAGAATTC GGTTTCGCGG GGGCTGTCGG TGTGTTGCGG
 501  TTCGGTNTGG AGGGTTTTGT CNCCGTTCGG GTNGAATGTG CTGACGATGC
 551  CTATTGCCAA TGCGCCGATG GCGGTGATAC AGATGAGCAA TACGGCGCGT
 601  ATCAGGAGTT TGGGGGTCAG CCTGAAGGGT TTGTTCNGTT TTTTTGCCAT
 651  TTTGATTGTG CTTTTGGGGT GTCGGGCAAT GCCGTCTGAA GGCGGTTCAG
 701  ACGGCATTGC CGAGTCAGCG TTGGACGTAG TTTNGGTAGA GGGTGATGAC
 751  TTTTTGTACG CCGACGGTGG TGCTGACTTT TTGGGTAATC TGCGCCTGTT
 801  CTTCGGGGGT GAGGATGCCC ATAACGTAGG TTACGTTGCC GTAGGTAACG
 851  ATTTTGACGC GCGCCTGTGT GGCGGGGCTG ATGCCCAACA GCGTGGCGCG
 901  GACTTTGGAT GTGTTCCAAG TGTCGCCGGC GATGTCGCCG GCAGTGCGCG
 951  GCAGGGAGGC GACGGTAATG TANTTGTACA CGCCTTCGGC GGCCTGTTCG
1001  GAACGTGCAA TCTGACCGAC GAACTGTTTC TCGCCTTCGG TGGCGACTTG
1051  TCCGAGCAGC AGCAGGTGGC GGTTGTAGCC GACAACGGAG ATTTGGGGCG
1101  TGTANCCTTT GGTTTGGTTG TTTTGGCGCA GATAGGAGCG GGCGGTGGTT
1151  TCGATACGCA GCGCCATTAC GTTGTCGTCG GTTNGCGCGC CGGTGGTTCG
1201  GCGGTCGACG GCGGATTTCG CGCCGACCGC CGCGCCGCCG ACGACTGCGC
1251  TGACGCAGCC GCCGAGGGCA AGGCTGAGGA CGGCGGCAGT CAGGGTGCGG
1301  ACGGTGTGCG GTTTGGGTTT CATCGGGTGC TTCCTTTCTT GGGCGTTTCA
1351  GACGGCATTG CTTTGCGCCA TGCCGTCTGA
```

This encodes a protein having amino acid sequence <SEQ ID 212>:

```
  1  MMXPXIMLPW IAGVPAVPGQ KRLSRXSLWG LGGXFFGVSG LVWFSLGVSX
 51  SLGVSXGCAC FSGVSFRGSG RGTFVGSTGV SLSVFSACAP ASSGCLSVXA
101  VSAGCGLTRX FXGAAGDGSP LPLSSVPSGC AGADEEAXXC SGWAASCPTT
151  PFGSQNSVSR GLSVCCGSVW RVLSPFGXNV LTMPIANAPM AVIQMSNTAR
201  IRSLGVSLKG LFXFFAILIV LLGCRAMPSE GGSDGIAESA LDVVXVEGDD
251  FLYADGGADF LGNLRLFFGG EDAHNVGYVA VGNDFDARLC GGADAQQRGA
301  DFGCVPSVAG DVAGSARQGG DGNVXVHAFG GLFGTCNLTD ELFLAFGGDL
351  SEQQQVAVVA DNGDLGRVXF GLVVLAQIGA GGGFDTQRHY VVVGXRAGGS
401  AVDGGFRADR RAADDCADAA AEGKAEDGGS QGADGVRFGF HRVLPFLGVS
451  DGIALRHAV*
```

ORF34a and ORF34-1 show 91.3% identity in 459 aa overlap:

```
                      10        20        30        40        50        60
orf34a.pep    MMXPXIMLPWIAGVPAVPGQKRLSRXSLWGLGGXFFGVSGLVWFSLGVSXSLGVSXGCAC
              || | |||||||||||||||:|||| ||||||| |||||||||||||||       ||||
orf34-1       MMMPFIMLPWIAGVPAVPGQNRLSRISLWGLGGVFFGVSGLVWFSLGVSL------GCAC
                      10        20        30        40        50

                      70        80        90       100       110       120
orf34a.pep    FSGVSFRGSGRGTFVGSTGVSLSVFSACAPASSGCLSVXAVSAGCGLTRXFXGAAGDGSP
              ||||||||||||||||||||||||||||:|||||||||||||||||||| | ||||||||
orf34-1       FSGVSFRGSGRGTFVGSTGVSLSVFSACVPASSGCLSVXAVSAGCGLTRFFLGAAGDGSP
                  60        70        80        90       100       110

                     130       140       150       160       170       180
orf34a.pep    LPLSSVPSGCAGADEEAXXCSGWAASCPTTPFGSQNSVSRGLSVCCGSVWRVLSPFGXNV
              ||||||||||||:|| |  |||||||||||||||||||||||||||||: ||||||| ||
orf34-1       LPLSSVPSGCAGSDEAAWWCSGWAASCPTTPFGSQNSVSRGLSVCCGSAXRVLSPFGLNV
                 120       130       140       150       160       170

                     190       200       210       220       230       240
orf34a.pep    LTMPIANAPMAVIQMSNTARIRSLGVSLKGLFXFFAILIVLLGCRAMPSEGGSDGIAESA
              |||||||||||:||||||||||||||||||||| ||||||||||||||||||||||||||
orf34-1       LTMPIANAPMAAIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA
                 180       190       200       210       220       230

                     250       260       270       280       290       300
orf34a.pep    LDVVXVEGDDFLYADGGADFLGNLRLFFGGEDAHNVGYVAVGNDFDARLCGGADAQQRGA
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf34-1       LDVVLVEGDDFLYADGGADFLGNLRLFFGGEDAHNVGYVAVGNDFDARLCGGADAQQRGA
                 240       250       260       270       280       290

                     310       320       330       340       350       360
orf34a.pep    DFGCVPSVAGDVAGSARQGGDGNVXVHAFGGLFGTCNLTDELFLAFGGDLSEQQQVAVVA
              |||||||||||||||||||||||: ||||||||||||||||||:||||||||||||||||
orf34-1       DFGCVPSVAGDVAGSARQGGDGNIVVHAFGGLFGTCNLTDELFFAFGGDLSEQQQVAVVA
                 300       310       320       330       340       350

                     370       380       390       400       410       420
orf34a.pep    DNGDLGRVXFGLVVLAQIGAGGGFDTQRHYVVVGXRAGGSAVDGGFRADRRAADDCADAA
              |:|||||| ||||||||||:||||||||| |||| ||||||||||||||  |:| |||||
orf34-1       DDGDLGRVAFGLVVLAQIGTGGGFDTQRHNVVVGLRAGGSAVDGGFRADGGASDYCADAA
                 360       370       380       390       400       410

                     430       440       450       460
orf34a.pep    AEGKAEDGGSQGADGVRFGFHRVLPFLGVSDGIALRHAVX
              |:||||:||:||||||||||||||||||||||||||||||
orf34-1       AKGKAENGGNQGADGVRFGFHRVLPFLGVSDGIALRHAVX
                 420       430       440       450
```

<u>Homology with a predicted ORF from *N.gonorrhoeae*</u>

ORF34 shows 77.6% identity over a 161aa overlap with a predicted ORF (ORF34.ng) from *N. gonorrhoeae:*

```
orf34.pep                 QKSLSRISLWGLGGVFFGVSGLVWFSLGVSXE------CAC   35
```

```
                                   || |||||||||:|||||||||||||||            |||
orf34ng     MMMPFIMLPWIAGVPAVPGQKRLSRISLWGLAGVFFGVSGLVWFSLGVSFSLGVSLGCAC   60

orf34.pep   FSGVSFRGSGRGTFVGSTGVSLSVFSACVXGVVRLPVGLSCV-----GRLXXLTRFFLGA   90
            ||||||||| |:||||||||||||||||        :||: | :      ||   |||||||
orf34ng     FSGVSFRGSGWGAFVGSTGVSLSVFSACVP----VPVNESAARAASEGR--GLTRFFLGA  114

orf34.pep   AGDVILLPLSSVPSGCAGSDEAAWWCSGWAASCPTTPFGSQNSVSRGLSVCCGSAXRVLS  150
            |||   ||||||||||||||||||||||||||||:|||||||||||||||||||: ||||
orf34ng     AGDGSPLPLSSVPSGCAGSDEAAWWCSGWAASCPTAPFGSQNSVSRGLSVCCGSVWRVLS  174

orf34.pep   S                                                            175

orf34ng     PFGLNVLTMPTANAPMAVIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSD  234
```

The complete length ORF34ng nucleotide sequence <SEQ ID 213> is:

```
   1  ATGATGATGC CGTTCATAAT GCTTCCTTGG ATTGCGGGTG TGCCTGCCGT
  51  GCCGGGTCAA AAGAGGTTGT CGAGAATCTC TTTATGGGGT TTGGCCGGCG
 101  TGTTTTTCGG GGTGTCCGGT TTGGTATGGT TTTCTTTGGG CGTTTCTTTT
 151  TCTTTGGGTG TTTCTTTGGG CTGCGCCTGT TTTTCGGGTG TTTCTTTTCG
 201  GGGTTCGGGA TGGGGGGCGT TTGTGGGCAG TACGGGGGTT TCTTTGAGTG
 251  TGTTTTCAGC TTGTGTTCCG GTGCCGGTTA ACGAATCGGC TGCCCGGGCC
 301  GCATCCGAAG GGCGCGGTTT gACCCGGTTT TTCTTGGGTG CGGCAGGGGA
 351  CGGCAGTCCG CTGCCGCTTT CTTCTGTGCC GTCCGGCTGT GCGGGTTCGG
 401  ATGAGGCGGC GTGGTGGTGT TCGGGTTGGG CGGCATCTTG TCCGACGGCG
 451  CCGTTTGGCA GCCAGAATTC GGTTTCGCGG GGGCTGTCGG TGTGTTGCGG
 501  TTCGGTTTGG AGGGTTTTGT CGCCGTTCGG GTTGAATGTG CTGACGATGC
 551  CTACTGCCAA TGCGCCGATG GCGGTGATAC AGATGAGCAA TACGGCGCGT
 601  ATCAGGAGTT TGGGGGTCAG CCTGAAGGGT TTGTTCGGTT TTTTTGCCAT
 651  TTTGATTGTG CTTTTGGGGT GTCGGGCAAT GCCGTCTGAA GGCGGTTCAG
 701  ACGGCATTGC CGAGTCAGCG TTGGACGTAG TTTTGGTAGA GGGTAATGAC
 751  TTTTTGTACG CCGAcggTGG TGCTGACTTT TTGGGTAATC TGCGCCTGTT
 801  CTTCGGGGGT GAGGATGCCC ATAACGTAGG TTACATTGCC GTAGGTAATG
 851  ATTTTGACGC GCGCCTGTGT AGCGGGGCTG ATGCCCAGCA GcgtgGCGCG
 901  GACTTTGGAC GTGTTCCAAG TGTCGCCGGC GATGTCGCCC GCAGTGCGCG
 951  GCAGGGAGGC GACGGTAATG TAGTTGTATA CGCCTTCGGC GGCCTGTTCG
1001  GAACGTGCAA TCTGACCGAC GAACTGTTTT TCGCCTTCGG TGGCGACTTG
1051  TCCGAGCAGC AGCAGGTGGC GGTTGTAGCC GACGACGGAG ATTTGGGGCG
1101  TGTAGCCTTT GGTTTGGTTG TTTTGGCGCA GGTAGGAACG GGCGGTGGTT
1151  TCGATACGCA ACGCCATAAC GTtgtCATCG GTTtgcgcgc CGGTGGTTcg
1201  gCGGTCGATG ACGGATTTTG CGCCGACGGC GGCCCCGCCG ACGACTGCGC
1251  TGAAGCAGCC GCCGAGGGCA AGGCTGAGGA CGGCGGCAAT CAGGGTGCGG
1301  ACGGTGTGTG GTTTGGGTTT CATCGGGGAC TTCCTTTCTT GGGCGTTTCA
1351  GACGGCATTG CTTTGCGCCA TGCCGTCTGA
```

This encodes a protein having amino acid sequence <SEQ ID 214>:

```
  1  MMMPFIMLPW IAGVPAVPGQ KRLSRISLWG LAGVFFGVSG LVWFSLGVSF
 51  SLGVSLGCAC FSGVSFRGSG WGAFVGSTGV SLSVFSACVP VPVNESAARA
101  ASEGRGLTRF FLGAAGDGSP LPLSSVPSGC AGSDEAAWWC SGWAASCPTA
151  PFGSQNSVSR GLSVCCGSVW RVLSPFGLNV LTMPTANAPM AVIQMSNTAR
201  IRSLGVSLKG LFGFFAILIV LLGCRAMPSE GGSDGIAESA LDVVLVEGND
251  FLYADGGADF LGNLRLFFGG EDAHNVGYIA VGNDFDARLC SGADAQQRGA
301  DFGRVPSVAG DVARSARQGG DGNVVVYAFG GLFGTCNLTD ELFFAFGGDL
351  SEQQQVAVVA DDGDLGRVAF GLVVLAQVGT GGGFDTQRHN VVIGLRAGGS
401  AVDDGFCADG GPADDCAEAA AEGKAEDGGN QGADGVWFGF HRGLPFLGVS
451  DGIALRHAV*
```

ORF34ng and ORF34-1 show 90.0% identity in 459 aa overlap:

```
                     10        20        30        40      4     50
orf34-1.pep  MMMPFIMLPWIAGVPAVPGQNRLSRISLWGLGGVFFGVSGLVWFSLGVS------LGCAC
             ||||||||||||||||||||:||||||||||:|||||||||||||||||      |||||
orf34ng      MMMPFIMLPWIAGVPAVPGQKRLSRISLWGLAGVFFGVSGLVWFSLGVSFSLGVSLGCAC
                     10        20        30        40        50        60

               60        70        80        90       100       110
orf34-1.pep  FSGVSFRGSGRGTFVGSTGVSLSVFSACVPASSGCLSVXAVSAGCGLTRFFLGAAGDGSP
             ||||||||| |:|||||||||||||||||:  :  :: |:| | ||||||||||||||||
orf34ng      FSGVSFRGSGWGAFVGSTGVSLSVFSACVPVPVNESAARAASEGRGLTRFFLGAAGDGSP
```

```
                          70        80        90       100       110       120

                     120       130       140       150       160       170
orf34-1.pep    LPLSSVPSGCAGSDEAAWWCSGWAASCPTTPFGSQNSVSRGLSVCCGSAXRVLSPFGLNV
               |||||||||||||||||||||||||||||:||||||||||||||||||: ||||||||||
orf34ng        LPLSSVPSGCAGSDEAAWWCSGWAASCPTAPFGSQNSVSRGLSVCCGSVWRVLSPFGLNV
                       130       140       150       160       170       180

                     180       190       200       210       220       230
orf34-1.pep    LTMPIANAPMAAIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA
               |||| ||||||:||||||||||||||||||||||||||||||||||||||||||||||||
orf34ng        LTMPTANAPMAVIQMSNTARIRSLGVSLKGLFGFFAILIVLLGCRAMPSEGGSDGIAESA
                       190       200       210       220       230       240

                     240       250       260       270       280       290
orf34-1.pep    LDVVLVEGDDFLYADGGADFLGNLRLFFGGEDAHNVGYVAVGNDFDARLCGGADAQQRGA
               ||||||||:|||||||||||||||||||||||||||||:|||||||||||:|||||||||
orf34ng        LDVVLVEGNDFLYADGGADFLGNLRLFFGGEDAHNVGYIAVGNDFDARLCSGADAQQRGA
                       250       260       270       280       290       300

                     300       310       320       330       340       350
orf34-1.pep    DFGCVPSVAGDVAGSARQGGDGNIVVHAFGGLFGTCNLTDELFFAFGGDLSEQQVAVVA
               ||| ||||||||| |||||||||:||:|||||||||||||||||||||||||||||||||
orf34ng        DFGRVPSVAGDVARSARQGGDGNVVVYAFGGLFGTCNLTDELFFAFGGDLSEQQVAVVA
                       310       320       330       340       350       360

                     360       370       380       390       400       410
orf34-1.pep    DDGDLGRVAFGLVVLAQIGTGGGFDTQRHNVVVGLRAGGSAVDGGFRADGGASDYCADAA
               |||||||||||||||||:||||||||||||||:|||||||||| || |||| :| ||:||
orf34ng        DDGDLGRVAFGLVVLAQVGTGGGFDTQRHNVVIGLRAGGSAVDDGFCADGGPADDCAEAA
                       370       380       390       400       410       420

                     420       430       440       450
orf34-1.pep    AKGKAENGGNQGADGVRFGFHRVLPFLGVSDGIALRHAVX
               |:||||:||||||||| ||||| ||||||||||||||||
orf34ng        AEGKAEDGGNQGADGVWFGFHRGLPFLGVSDGIALRHAVX
                       430       440       450       460
```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 26

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 215>:

```
  1  ATGAAAACCT TCTTCAAAAC CCTTTCCGCC GCCGCACTCG CGCTCATCCT
 51  CGCCGCCTGC GGATT.CAAA AAGACAGCGC GCCCGCCGCA TCCGCTTCTG
101  CCGCCGCCGA CAACGGCGCG GCGTAAAAAA GAAATCGTCT TCGGCACGAC
151  CGTCGGCGAC TTCGGCGATA TGGTCAAAGA ACAAATCCAA GCCGAGCTGG
201  AGAAAAAAGG CTACACCGTC AAACTGGTCG AGTTTACCGA CTATGTACGC
251  CCGAATCTGG CATTGGCTGA GGGCGAGTTG
```

This corresponds to the amino acid sequence <SEQ ID 216; ORF4>:

```
  1  MKTFFKTLSA AALALILAAC G.QKDSAPAA SASAAADNGA AKKEIVFGTT
 51  VGDFGDMVKE QIQAELEKKG YTVKLVEFTD YVRPNLALAE GEL
```

Further sequence analysis revealed the complete nucleotide sequence <SEQ ID 217>:

```
  1  ATGAAAACCT TCTTCAAAAC CCTTTCCGCC GCCGCACTCG CGCTCATCCT
 51  CGCCGCCTGC GGCGGTCAAA AAGACAGCGC GCCCGCCGCA TCCGCTTCTG
101  CCGCCGCCGA CAACGGCGCG GCGAAAAAAG AAATCGTCTT CGGCACGACC
151  GTCGGCGACT TCGGCGATAT GGTCAAAGAA CAAATCCAAG CCGAGCTGGA
201  GAAAAAAGGC TACACCGTCA AACTGGTCGA GTTTACCGAC TATGTACGCC
```

```
251  CGAATCTGGC ATTGGCTGAG GGCGAGTTGG ACATCAACGT CTTCCAACAC
301  AAACCCTATC TTGACGACTT CAAAAAAGAA CACAATCTGG ACATCACCGA
351  AGTCTTCCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
401  AATCGCTGGA AGAAGTCAAA GACGGCAGCA CCGTATCCGC GCCCAACGAC
451  CCGTCCAACT TCGCCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTGGAT
501  CAAACTCAAA GACGGCATCA ATCCGTTGAC CGCATCCAAA GCGGACATCG
551  CCGAGAACCT GAAAAACATC AAAATCGTCG AGCTTGAAGC CGCGCAACTG
601  CCGCGTAGCC GCGCCGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC
651  CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
701  TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
751  TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
801  CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGGAATGAAG
851  GCGCAGCCAA ATAA
```

This corresponds to the amino acid sequence <SEQ ID 218; ORF4-1>:

```
  1  MKTFFKTLSA AALALILAAC GGQKDSAPAA SASAAADNGA AKKEIVFGTT
 51  VGDFGDMVKE QIQAELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFQH
101  KPYLDDFKKE HNLDITEVFQ VPTAPLGLYP GKLKSLEEVK DGSTVSAPND
151  PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLKNI KIVELEAAQL
201  PRSRADVDFA VVNGNYAISS GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ
251  WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF4 shows 93.5% identity over a 93aa overlap with an ORF (ORF4a) from strain A of *N. meningitidis:*

```
                        10        20        30        40        50        59
orf4.pep        MKTFFKTLSAAALALILAACG-QKDSAPAASASAAADNGAAKKEIVFGTTVGDFGDMVKE
                |||||||||||||||||||| |||||||||||||||||| |||||||||||||||||
orf4a           MKTFFKTLSAAALALILAACGGQKDSAPAASASAAADNGAAXKEIVFGTTVGDFGDMVKE
                        10        20        30        40        50        60

               60        70        80        90
orf4.pep        QIQAELEKKGYTVKLVEFTDYVRPNLALAEGEL
                 || ||||||||||||| ||||| |||||||||
orf4a           XIQPELEKKGYTVKLVEXTDYVRXNLALAEGELDINVXQHXXYLDDXKKXHNLDITXVXQ
                        70        80        90       100       110       120

orf4a           VPTAPLGLYPGKLKSLXXVKXGSTVSAPNDPXXFXRVLVMLDELGXIKLKDXIXXXXXXX
                       130       140       150       160       170       180
```

The complete length ORF4a nucleotide sequence <SEQ ID 219> is:

```
  1  ATGAAAACCT TCTTCAAAAC CCTTTCCGCC GCCGCACTCG CGCTCATCCT
 51  CGCCGCCTGC GGCGGTCAAA AAGATAGCGC GCCCGCCGCA TCCGCTTCTG
101  CCGCCGCCGA CAACGGCGCG GCGAANAAAG AAATCGTCTT CGGCACGACC
151  GTCGGCGACT TCGGCGATAT GGTCAAAGAA CANATCCAAC CCGAGCTGGA
201  GAAAAAAGGC TACACCGTCA AACTGGTCGA GTNTACCGAC TATGTGCGCN
251  CGAATCTGGC ATTGGCTGAG GGCGAGTTGG ACATCAACGT CTTNCAACAC
301  ANACNCTATC TTGACGACTN CAAAAAANAA CACAATCTGG ACATCACCNN
351  AGTCTTNCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
401  AATCGCTGGA NNAAGTCAAA GANGGCAGCA CCGTATCCGC GCCCAACGAC
451  CCGTNNNACT TCGNCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTNGAT
501  CAAACTCAAA GACNGCATCA NNNNGNNGNN NNNANCNANA NNNGANANNN
551  NNNNANNNNT NNNNNNNNNN NNNNNCNNCG NNNNNNNANN NNNNNNNNNN
601  NCGNNTNNNN NNGCNNNNNT NNANNNTNNN NNCNNCNNNN NNNNNTNNNN
651  NANNANNAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
701  TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
751  TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
801  CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGGAATGAAG
851  GCGCAGCCAA ATAA
```

This is predicted to encode a protein having amino acid sequence <SEQ ID 220>:

```
  1  MKTFFKTLSA AALALILAAC GGQKDSAPAA SASAAADNGA AXKEIVFGTT
```

```
 51  VGDFGDMVKE XIQPELEKKG YTVKLVEXTD YVRXNLALAE GELDINVXQH
101  XXYLDDXKKX HNLDITXVXQ VPTAPLGLYP GKLKSLXXVK XGSTVSAPND
151  PXXFXRVLVM LDELGXIKLK DXIXXXXXXX XXXXXXXXXX XXXXXXXXXX
201  XXXXAXXXXX XXXXXXXXXS GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ
251  WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK*
```

A leader peptide is underlined.

Further analysis of these strain A sequences revealed the complete DNA sequence <SEQ ID 221>:

```
  1  ATGAAAACCT TCTTCAAAAC CCTTTCCGCC GCCGCACTCG CGCTCATCCT
 51  CGCCGCCTGC GGCGGTCAAA AAGATAGCGC GCCCGCCGCA TCCGCTTCTG
101  CCGCCGCCGA CAACGGCGCG GCGAAAAAAG AAATCGTCTT CGGCACGACC
151  GTCGGCGACT TCGGCGATAT GGTCAAAGAA CAAATCCAAC CCGAGCTGGA
201  GAAAAAAGGC TACACCGTCA AACTGGTCGA GTTTACCGAC TATGTGCGCC
251  CGAATCTGGC ATTGGCTGAG GGCGAGTTGG ACATCAACGT CTTCCAACAC
301  AAACCCTATC TTGACGACTT CAAAAAAGAA CACAATCTGG ACATCACCGA
351  AGTCTTCCAA GTGCCGACCG CGCCTTTGGG ACTGTACCCG GGCAAGCTGA
401  AATCGCTGGA AGAAGTCAAA GACGGCAGCA CCGTATCCGC GCCCAACGAC
451  CCGTCCAACT TCGCCCGCGT CTTGGTGATG CTCGACGAAC TGGGTTGGAT
501  CAAACTCAAA GACGGCATCA ATCCGCTGAC CGCATCCAAA GCGGACATTG
551  CCGAAAACCT GAAAAACATC AAAATCGTCG AGCTTGAAGC CGCGCAACTG
601  CCGCGTAGCC GCGCCGACGT GGATTTTGCC GTCGTCAACG GCAACTACGC
651  CATAAGCAGC GGCATGAAGC TGACCGAAGC CCTGTTCCAA GAACCGAGCT
701  TTGCCTATGT CAACTGGTCT GCCGTCAAAA CCGCCGACAA AGACAGCCAA
751  TGGCTTAAAG ACGTAACCGA GGCCTATAAC TCCGACGCGT TCAAAGCCTA
801  CGCGCACAAA CGCTTCGAGG GCTACAAATC CCCTGCCGCA TGGAATGAAG
851  GCGCAGCCAA ATAA
```

This encodes a protein having amino acid sequence <SEQ ID 222; ORF4a-1>:

```
  1  MKTFFKTLSA AALALILAAC GGQKDSAPAA SASAAADNGA AKKEIVFGTT
 51  VGDFGDMVKE QIQPELEKKG YTVKLVEFTD YVRPNLALAE GELDINVFQH
101  KPYLDDFKKE HNLDITEVFQ VPTAPLGLYP GKLKSLEEVK DGSTVSAPND
151  PSNFARVLVM LDELGWIKLK DGINPLTASK ADIAENLKNI KIVELEAAQL
201  PRSRADVDFA VVNGNYAISS GMKLTEALFQ EPSFAYVNWS AVKTADKDSQ
251  WLKDVTEAYN SDAFKAYAHK RFEGYKSPAA WNEGAAK*
```

ORF4a-1 and ORF4-1 show 99.7% identity in 287 aa overlap:

```
                    10        20        30        40        50        60
orf4a-1    MKTFFKTLSAAALALILAACGGQKDSAPAASASAAADNGAAKKEIVFGTTVGDFGDMVKE
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf4-1     MKTFFKTLSAAALALILAACGGQKDSAPAASASAAADNGAAKKEIVFGTTVGDFGDMVKE
                    10        20        30        40        50        60

                    70        80        90       100       110       120
orf4a-1    QIQPELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEVFQ
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf4-1     QIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEVFQ
                    70        80        90       100       110       120

                   130       140       150       160       170       180
orf4a-1    VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf4-1     VPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTASK
                   130       140       150       160       170       180

                   190       200       210       220       230       240
orf4a-1    ADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf4-1     ADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNWS
                   190       200       210       220       230       240

                   250       260       270       280
orf4a-1    AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAKX
           ||||||||||||||||||||||||||||||||||||||||||||||||
orf4-1     AVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAKX
                   250       260       270       280
```

Homology with an outer membrane protein of *Pasteurella haemolitica* (accession q08869).

ORF4 and this outer membrane protein show 33% aa identity in 91aa overlap:

```
                                                     10        20
lip2.pasha                                   MNFKKLLGVALVSALALTACKDEKAQAP----
                                               || | ::||  || |:||  :|:  |
ORF4             VXTPNPDGRTPCPSFLFETATTSGENMKTFFKTLSAAAL--ALILAACGFKKTARPPHPL
                       110       120       130       140         150

                  30        40        50        60        70        80
lip2.pasha       -ATTAKTENKAPLKVGVMTGPEAQMTEVAVKIAKEKYGLDVELVQFTEYTQPNAALHSKD
                     : :: |  :  |:  |  ::|::  ::   || |  |:||:||:|::|| ||   :
ORF4             LPPPTTARRKKEIVFGTTVGDFGDMVKEQIQAELEKKGYTVKLVEFTDYVRPNLALAEGE
                160       170       180       190       200       210

                  90        100       110       120       130       140
lip2.pasha       LDANAFQTVPYLEQEVKDRGYKLAIIGNTLVWPIAAYSKKIKNISELKDGATVAIPNNAS
                 |
ORF4             L.....
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF4 shows 93.6% identity over a 94aa overlap with a predicted ORF (ORF4.ng) from *N. gonorrhoeae:*

```
                                                       10        20        30
orf4nm.pep                                     MKTFFKTLSAAALALILAACGXQKDSAPAA
                                               |||||||||:|:||||||||| ||||||||
orf4ng           RANAVXTPNPDGRTPCLSFLFETATTSGENMKTFFKTLSTASLALILAACGGQKDSAPAA
                        200       210       220       230       240       250

                          40        50        60        70        80        89
orf4nm.pep       SASA-AADNGAAKKEIVFGTTVGDFGDMVKEQIQAELEKKGYTVKLVEFTDYVRPNLALA
                 ||:| :||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf4ng           SAAAPSADNGAAKKEIVFGTTVGDFGDMVKEQIQAELEKKGYTVKLVEFTDYVRPNLALA
                        260       270       280       290       300       310

                90
orf4nm.pep       EGEL
                 ||||
orf4ng           EGELDINVFQHKPYLDDFKKEHNLDITEAFQVPTAPLGLYPGKLKSLEEVKDGSTVSAPN
                        320       330       340       350       360       370
```

The complete length ORF4ng nucleotide sequence <SEQ ID 223> was predicted to encode a protein having amino acid sequence <SEQ ID 224>:

```
  1 MKTFFKTLST ASLALILAAC GGQKDSAPAA SAAAPSADNG AAKKEIVFGT
 51 TVGDFGDMVK EQIQAELEKK GYTVKLVEFT DYVRPNLALA EGELDINVFQ
101 HKPYLDDFKK EHNLDITEAF QVPTAPLGLY PGKLKSLEEV KDGSTVSAPN
151 DPSNFARALV MLNELGWIKL KDGINPLTAS KADIAENLKN IKIVELEAAQ
201 LPRSRADVDF AVVNGNYAIS SGMKLTEALF QEPSFAYVNW SAVKTADKDS
251 QWLKDVTEAY NSDAFKAYAH KRFEGYKYPA AWNEGAAK*
```

Further analysis revealed the complete length ORF4ng DNA sequence <SEQ ID 225> to be:

```
  1 atgAAAACCT TCTTCAAAAC cctttccgcc gccgcaCTCG CGCTCATCCT
 51 CGCAGCCTGc ggCggtcaAA AAGACAGCGC GCCCgcagcc tctgcCGCCG
101 CCCCTTCTGC CGATAACGGc gCgGCGAAAA AAGAAAtcgt ctTCGGCACG
151 Accgtgggcg acttcggcgA TAtggTCAAA GAACAAATCC AagcCGAgct
201 gGAGAAAAAA GgctACACCg tcAAattggt cgaatttacc gactatgtGC
251 gCCCGAATCT GGCATTGGCG GAGGGCGAGT TGGACATCAA CGTCTTCCAA
301 CACAAACCCT ATCTTGACGA TTTCAAAAAA GAACACAACC TGGACATCAC
351 CGAAGCCTTC CAAGTGCCGA CCGCGCCTTT GGGACTGTAT CCGGGCAAAC
401 TGAAATCGCT GGAAGAAGTC AAAGACGGCA GCACCGTATC CGCGCCCAac
451 gACccgTCCA ACTTCGCACG CGCCTTGGTG ATGCTGAACG AACTGGGTTG
501 GATCAAACTC AAAGACGGCA TCAATCCGCT GACCGCATCC AAAGCCGACA
551 TCGCGGAAAA CCTGAAAAAC ATCAAAATCG TCGAGCTTGA AGCCGCACAA
```

```
601   CTGCCGCGCA GCCGCGCCGA CGTGGATTTT GCCGTCGTCA ACGGCAACTA
651   CGCCATAAGC AGCGGCATGA AGCTGACCGA AGCCCTGTTC CAAGAGCCGA
701   GCTTTGCCTA TGTCAACTGG TCTGCCgtcA AAACCGCCGA CAAAGACAGC
751   CAATGGCTTA AAGACGTAAC CGAGGCCTAT AACTCCGACG CGTTCAAAGC
801   CTACGCGCAC AAACGCTTCG AGGGCTACAA ATACCCTGCC GCATGGAATG
851   AAGGCGCAGC CAAATAA
```

This encodes a protein having amino acid sequence <SEQ ID 226; ORF4ng-1>:

```
  1   MKTFFKTLSA AALALILAAC GGQKDSAPAA SAAAPSADNG AAKKEIVFGT
 51   TVGDFGDMVK EQIQAELEKK GYTVKLVEFT DYVRPNLALA EGELDINVFQ
101   HKPYLDDFKK EHNLDITEAF QVPTAPLGLY PGKLKSLEEV KDGSTVSAPN
151   DPSNFARALV MLNELGWIKL KDGINPLTAS KADIAENLKN IKIVELEAAQ
201   LPRSRADVDF AVVNGNYAIS SGMKLTEALF QEPSFAYVNW SAVKTADKDS
251   QWLKDVTEAY NSDAFKAYAH KRFEGYKYPA AWNEGAAK*
```

This shows 97.6% identity in 288 aa overlap with ORF4-1:

```
                      10        20        30        40        50        59
orf4-1.pep    MKTFFKTLSAAALALILAACGGQKDSAPAASASA-AADNGAAKKEIVFGTTVGDFGDMVK
              ||||||||||||||||||||||||||||||||:| :||||||||||||||||||||||||
orf4ng-1      MKTFFKTLSAAALALILAACGGQKDSAPAASAAAPSADNGAAKKEIVFGTTVGDFGDMVK
                      10        20        30        40        50        60

                    60        70        80        90       100       110       119
orf4-1.pep    EQIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEVF
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|
orf4ng-1      EQIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITEAF
                      70        80        90       100       110       120

                   120       130       140       150       160       170       179
orf4-1.pep    QVPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARVLVMLDELGWIKLKDGINPLTAS
              |||||||||||||||||||||||||||||||||||||:||||:|||||||||||||||||
orf4ng-1      QVPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARALVMLNELGWIKLKDGINPLTAS
                     130       140       150       160       170       180

                   180       190       200       210       220       230       239
orf4-1.pep    KADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf4ng-1      KADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTEALFQEPSFAYVNW
                     190       200       210       220       230       240

                   240       250       260       270       280
orf4-1.pep    SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKSPAAWNEGAAKX
              ||||||||||||||||||||||||||||||||||||| |||||||||||
orf4ng-1      SAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKYPAAWNEGAAKX
                     250       260       270       280
```

In addition, ORF4ng-1 shows significant homology with an outer membrane protein from the database:

```
ID    LIP2_PASHA      STANDARD;        PRT;    276 AA.
AC    Q08869;
DT    01-NOV-1995 (REL. 32, CREATED)
DT    01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT    01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE    28.2 KD OUTER MEMBRANE PROTEIN PRECURSOR. . . .
SCORES       Init1:   279 Initn:   416 Opt:   494
Smith-Waterman score: 494;    36.0% identity in 275 aa overlap

                       10          20        30        40        50
orf4ng-1.pep    MKTFFKTLSAAAL--ALILAACGGQKDSAPAASAAAPSADNGAAKKEIVFGTTVGDFGDM
                    || | ::||  || |:||  :| :|||::|   :::| |  |    |: :|  ::|
lip2_pasha        MNFKKLLGVALVSALALTACKDEKAQAPATTA---KTENKAPLK---VGVMTGPEAQM
                         10        20        30           40            50

                60        70        80        90       100       110
orf4ng-1.pep    VKEQIQAELEKKGYTVKLVEFTDYVRPNLALAEGELDINVFQHKPYLDDFKKEHNLDITE
                ::  ::   || |  |:||:||:|::|| ||   :|| |:||  |||::  |:::  ::
```

```
lip2_pasha       TEVAVKIAKEKYGLDVELVQFTEYTQPNAALHSKDLDANAFQTVPYLEQEVKDRGYKLAI
                       60        70        80        90       100       110

                120       130       140       150       160       170
orf4ng-1.pep     AFQVPTAPLGLYPGKLKSLEEVKDGSTVSAPNDPSNFARALVMLNELGWIKLKDGINPLT
                   :: : |:: |  |:|:: |:|||:||: ||: || ||||::|:  | :||||  | :
lip2_pasha       IGNTLVWPIAAYSKKIKNISELKDGATVAIPNNASNTARALLLLQAHGLLKLKDPKN-VF
                      120       130       140       150       160       170

                180       190       200       210       220         230
orf4ng-1.pep     ASKADIAENLKNIKIVELEAAQLPRSRADVDFAVVNGNYAISSGMKLTE--ALFQEPSFA
                 |:: || || ||||||: ::: | |   ||::||:|::|| ::|::  :   : :  :
lip2_pasha       ATENDIIENPKNIKIVQADTSLLTRMLDDVELAVINNTYAGQAGLSPDKDGIIVESKDSP
                       180       190       200       210       220       230

                  240       250       260       270       280       289
orf4ng-1.pep     YVNWSAVKTADKDSQWLKDVTEAYNSDAFKAYAHKRFEGYKYPAAWNEGAAKX
                 |||  : :  :||:  |:  :::::::     | | |:|
lip2_pasha       YVNLVVSREDNKDDPRLQTFVKSFQTEEVFQEALKLFNGGVVKGW
                       240       250       260       270
```

Based on this analysis, including the homology with the outer membrane protein of *Pasteurella haemolitica*, and on the presence of a putative prokaryotic membrane lipoprotein lipid attachment site in the gonococcal protein, it was predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

ORF4-1 (30kDa) was cloned in pET and pGex vectors and expressed in *E.coli*, as described above. The products of protein expression and purification were analyzed by SDS-PAGE. Figures 8A and 8B show, repsectively, the results of affinity purification of the His-fusion and GST-fusion proteins. Purified His-fusion protein was used to immunise mice, whose sera were used for ELISA (positive result), Western blot (Figure 8C), FACS analysis (Figure 8D), and a bactericidal assay (Figure 8E). These experiments confirm that ORF4-1 is a surface-exposed protein, and that it is a useful immunogen.

Figure 8F shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF4-1.

Example 27

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 227>:

```
  1 CCTCGTCGTC CTCGGCATGC TCCAGTTTCA AGGGGCGATT TACTCCAAGG
 51 CGGTGGAACG TATGCTCGGC ACGGTCATCG GGCTGGGCGC GGGTTTGGGC
101 GTTTTATGGC TGAACCAGCA TTATTTCCAC GGCAACCTCC TCTTCTACCT
151 CACCGTCGGC ACGGCAAGCG CACTGGCCGG CTGGGCGGCG GTCGGCAAAA
201 ACGGCTACGT CCCTmTGCTG GCAGGGCTGA CGATGTGTAT GCTCATCGGC
251 GACAACGGCA GCGAATGGCT CGACAGCGGA CTCATGCGCG CCATGAACGT
301 CCTCATCGGC GyGGCCATCG CCATCGCCGC CGCCAAACTG CTGCCGCTGA
351 AATCCACACT GATGTGGCGT TTCATGCTTG CCGACAACCT GGCCGACTGC
401 AGCAAAATGA TTGCCGAAAT CAGCAACGGC AGGCGCATGA CCCGCGAACG
451 CCTCGAGGAG AACATGGCGA AAATGCGCCA AATCAACGCA CGCATGGTCA
501 AAAGCCGCAG CCATCTCGCC GCCACATCGG GCGAAAGCTG CATCAGCCCC
551 GCCATGATGG AAGCCATGCA GCACGCCCAC CGTAAAATCG TCAACACCAC
601 CGAGCTGCTC CTGACCACCG CCGCCAAGCT GCAATCTCCC AAACTCAACG
```

```
651  GCAGCGAAAT CCGGCTGCTT GACCGCCACT TCACACTGCT CCAAAC....
701  ............................ GC AGACACGCCC GCCGCATCCG
751  CATCGACACC GCCATCAACC CCGAACTGGA AGCCCTCGCC GAACACCTCC
801  ACTACCAATG GCAGGGCTTC CTCTGGCTCA GCACCGATAT GCGTCAGGAA
851  ATTTCCGCCC TCGTCATCCT GCTGCAACGC ACCCGCCGCA AATGGCTGGA
901  TGCCCACGAA CGCCAACACC TGCGCCAAAG CCTGCTTGA
```

This corresponds to the amino acid sequence <SEQ ID 228; ORF8>:

```
  1  ......PRRP RHAPVSRGDL LQGGGTYARH GHRAGRGFGR FMAEPALFPR
 51  QPPLLPHRRH GKRTGRLGGG RQKRLRPXAG RADDVYAHRR QRQRMARQRT
101  HARHERPHRR GHRHRRRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ
151  AHDPRTPRGE HGENAPNQRT HGQKPQPSRR HIGRKLHQPR HDGSHAARPP
201  XNRQHHRAAP DHRRQAAISQ TQRQRNPAAX PPLHTAPN.. .........Q
251  TRPPHPHRHR HQPRTGSPRR TPPLPMAGLP LAQHRYASGN FRPRHPAATH
301  PPQMAGCPRT PTPAPKPA*
```

Computer analysis of this amino acid sequence gave the following results:

Sequence motifs

ORF8 is proline-rich and has a distribution of proline residues consistent with a surface localization. Furthermore the presence of an RGD motif may indicate a possible role in bacterial adhesion events.

Homology with a predicted ORF from *N.gonorrhoeae*

ORF8 shows 86.5% identity over a 312aa overlap with a predicted ORF (ORF8.ng) from *N. gonorrhoeae:*

```
orf8ng      1 MDRDDRLRRPRHAPVPRRDLLQRGGTYARYGHRAGRGFGRFMAEPALFPR 50
                    |||||||| | |||| ||||||:||||||||||||||||||||
orf8.pep    1 ......PRRPRHAPVSRGDLLQGGGTYARHGHRAGRGFGRFMAEPALFPR 44

orf8ng     51 QPPLLPDHRHGKRTGRLGGGRQKRLRPYVGGADDVHAHRRQRQRMARQRP 100
              ||||||  |||||||||||||||||||   | ||||:|||||||||||||
orf8.pep   45 QPPLLPHRRHGKRTGRLGGGRQKRLRPXAGRADDVYAHRRQRQRMARQRT 94

orf8ng    101 DARDERPHRRRHRHCRRQTAAAEIHTDVAFHACRQPGRLQQNDCRNQQRQ 150
               || |||||| ||| |||||||||||||||||||||||| ||||||||||
orf8.pep   95 HARHERPHRRGHRHRRRQTAAAEIHTDVAFHACRQPGRMQQNDCRNQQRQ 144

orf8ng    151 AYDARTFGAEYGQNAPNQRTHGQKPQPPRRHIGRKPHQPLHDGSHAARPP 200
              |:| ||   |:|:|||||||||||||| ||||||| ||| ||||||||||
orf8.pep  145 AHDPRTPRGEHGENAPNQRTHGQKPQPSRRHIGRKLHQPRHDGSHAARPP 194

orf8ng    201 QNRQHHRAAPDHRRQAAISQTQRQRNPAARPPLHTAPNRPATNRRPHQRQ 250
               |||||||||||||||||||||||||||| ||||||||           |
orf8.pep  195 XNRQHHRAAPDHRRQAAISQTQRQRNPAAXPPLHTAPN...........Q 244

orf8ng    251 TRPPHPHRHRHQPRTGSPRRTPPLPMAGFPLAQHQYASGNFRPRHPPATH 300
              |||||||||||||||||||||||||||| |||||.|||||||||||| |||
orf8.pep  245 TRPPHPHRHRHQPRTGSPRRTPPLPMAGLPLAQHRYASGNFRPRHPAATH 294

orf8ng    301 PPQMAGCPRTPTPAPKPA* 319
              ||||||||||||||||||
orf8.pep  295 PPQMAGCPRTPTPAPKPA* 313
```

The complete length ORF8ng nucleotide sequence <SEQ ID 229> is predicted to encode a protein having amino acid sequence <SEQ ID 230>:

```
  1  MDRDDRLRRP RHAPVPRRDL LQRGGTYARY GHRAGRGFGR FMAEPALFPR
 51  QPPLLPDHRH GKRTGRLGGG RQKRLRPYVG GADDVHAHRR QRQRMARQRP
101  DARDERPHRR RHRHCRRQTA AAEIHTDVAF HACRQPGRLQ QNDCRNQQRQ
151  AYDARTFGAE YGQNAPNQRT HGQKPQPPRR HIGRKPHQPL HDGSHAARPP
```

```
201  QNRQHHRAAP DHRRQAAISQ TQRQRNPAAR PPLHTAPNRP ATNRRPHQRQ
251  TRPPHPHRHR HQPRTGSPRR TPPLPMAGFP LAQHQYASGN FRPRHPPATH
301  PPQMAGCPRT PTPAPKPA*
```

Based on the sequence motifs in these proteins, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 28

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 231>:

```
  1  ..GAAATCAGCC TGCGGTCCGA CNACAGGCCG GTTTCCGTGN CGAAGCGGCG
 51    GGATTCGGAA CGTTTTCTGC TGTTGGACGG CGGCAACAGC CGGCTCAAGT
101    GGGCGTGGGT GGAAAACGGC ACGTTCGCAA CCGTCGGTAG CGCGCCGTAC
151    CGCGATTTGT CGCCTTTGGG CGCGGAGTGG GCGGAAAAGG CGGATGGAAA
201    TGTCCGCATC GTCGGTTGCG CTGTGTGCGG AGAATTCAAA AAGGCACAAG
251    TGCAGGAACA GCTCGCCCGA AAAATCGAGT GGCTGCCGTC TTCCGCACAG
301    GCTTT.GGCA TACGCAACCA CTACCGCCAC CCCGAAGAAC ACGGTTCCGA
351    CCGCTGGTTC AACGCCTTGG GCAGCCGCCG CTTCAGCCGC AACGCCTGCG
401    TCGTCGTCAG TTGCGGCACG GCGGTAACGG TTGACGCGCT CACCGATGAC
451    GGACATTATC TCGGAGA.GG AACCATCATG CCCGGTTTCC ACCTGATGAA
501    AGAATCGCTC GCCGTCCGAA CCGCCAACCT CAACCGGCAC GCCGGTAAGC
551    GTTATCCTTT CCCGACCGG..
```

This corresponds to the amino acid sequence <SEQ ID 232; ORF61>:

```
  1  ..EISLRSDXRP VSVXKRRDSE RFLLLDGGNS RLKWAWVENG TFATVGSAPY
 51    RDLSPLGAEW AEKADGNVRI VGCAVCGEFK KAQVQEQLAR KIEWLPSSAQ
101    AXGIRNHYRH PEEHGSDRWF NALGSRRFSR NACVVVSCGT AVTVDALTDD
151    GHYLGXGTIM PGFHLMKESL AVRTANLNRH AGKRYPFPT..
```

Further work revealed the complete nucleotide sequence <SEQ ID 233>:

```
   1  ATGACGGTTT TGAAGCTTTC GCACTGGCGG GTGTTGGCGG AGCTTGCCGA
  51  CGGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC
 101  CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA CATACGCGGG
 151  CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CATTGGCGGT
 201  TTTCGATGCC GAAGGTTTGC GCGAGCTGGG GGAAAGGTCG GGTTTTCAGA
 251  CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
 301  GCGCGGATTG CGCCGGACAA GGCGCACAAA ACCATATGCG TGACCCACCT
 351  GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
 401  GCGAGTGTCT GATGTTCAGT TTTGGCTGGG TGTTTGACCG GCCGCAGTAT
 451  GAGTTGGGTT CGCTGTCGCC TGTTGCGGCA GTGGCGTGTC GGCGCGCCTT
 501  GTCGCGTTTA GGTTTGGATG TGCAGATTAA GTGGCCCAAT GATTTGGTTG
 551  TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACGGT CAGGACGGGC
 601  GGCAAAACGG TTGCCGTGGT CGGTATCGGC ATCAATTTTG TCCTGCCCAA
 651  GGAAGTAGAA AATGCCGCTT CCGTGCAATC GCTGTTTCAG ACGGCATCGC
 701  GGCGGGGCAA TGCCGATGCC GCCGTGCTGC TGGAAACGCT GTTGGTGGAA
 751  CTGGACGCGG TGTTGTTGCA ATATGCGCGG GACGGATTTG CGCCTTTTGT
 801  GGCGGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
 851  TGCGCGACGG CGAAACCGTG TTCGAAGGCA CGGTTAAAGG CGTGGACGGA
 901  CAAGGCGTTT TGCACTTGGA AACGGCAGAG GGCAAACAGA CGGTCGTCAG
 951  CGGCGAAATC AGCCTGCGGT CCGACGACAG GCCGGTTTCC GTGCCGAAGC
1001  GGCGGGATTC GGAACGTTTT CTGCTGTTGG ACGGCGGCAA CAGCCGGCTC
1051  AAGTGGGCGT GGGTGGAAAA CGGCACGTTC GCAACCGTCG GTAGCGCGCC
1101  GTACCGCGAT TTGTCGCCTT TGGGCGCGGA GTGGGCGGAA AAGGCGGATG
1151  GAAATGTCCG CATCGTCGGT TGCGCTGTGT GCGGAGAATT CAAAAAGGCA
1201  CAAGTGCAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC
1251  ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT
1301  CCGACCGCTG GTTCAACGCC TTGGGCAGCC GCCGCTTCAG CCGCAACGCC
1351  TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
1401  TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGGTTTC CACCTGATGA
1451  AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGGCA CGCCGGTAAG
1501  CGTTATCCTT TCCCGACCAC AACGGGCAAT GCCGTCGCCA GCGGCATGAT
1551  GGATGCGGTT TGCGGCTCGG TTATGATGAT GCACGGGCGT TTGAAAGAAA
1601  AAACCGGGGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG
```

```
1651  GCAAAAGTTG CCGAAGCCCT GCCGCCTGCA TTTTTGGCGG AAAATACCGT
1701  GCGCGTGGCG GACAACCTCG TCATTTACGG GTTGTTGAAC ATGATTGCCG
1751  CCGAAGGCAG GGAATATGAA CATATTTAA
```

This corresponds to the amino acid sequence <SEQ ID 234; ORF61-1>:

```
  1  MTVLKLSHWR VLAELADGLP QHVSQLARMA DMKPQQLNGF WQQMPAHIRG
 51  LLRQHDGYWR LVRPLAVFDA EGLRELGERS GFQTALKHEC ASSNDEILEL
101  ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWVFDRPQY
151  ELGSLSPVAA VACRRALSRL GLDVQIKWPN DLVVGRDKLG GILIETVRTG
201  GKTVAVVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLVE
251  LDAVLLQYAR DGFAPFVAEY QAANRDHGKA VLLLRDGETV FEGTVKGVDG
301  QGVLHLETAE GKQTVVSGEI SLRSDDRPVS VPKRRDSERF LLLDGGNSRL
351  KWAWVENGTF ATVGSAPYRD LSPLGAEWAE KADGNVRIVG CAVCGEFKKA
401  QVQEQLARKI EWLPSSAQAL GIRNHYRHPE EHGSDRWFNA LGSRRFSRNA
451  CVVVSCGTAV TVDALTDDGH YLGGTIMPGF HLMKESLAVR TANLNRHAGK
501  RYPFPTTTGN AVASGMMDAV CGSVMMMHGR LKEKTGAGKP VDVIITGGGA
551  AKVAEALPPA FLAENTVRVA DNLVIYGLLN MIAAEGREYE HI*
```

Figure 9 shows plots of hydrophilicity, antigenic index, and AMPHI regions for ORF61-1. Further computer analysis of this amino acid sequence gave the following results:

<u>Homology with the baf protein of *B. pertussis* (accession number U12020).</u>

ORF61 and baf protein show 33% aa identity in 166aa overlap:

```
orf61  23  LLLDGGNSRLKWAWVE-NGTFATVGSAPYR----DLSPLGAEWAEKADGNVRIVGCAVCG 77
           +L+D GNSRLK  W + +    A    AP       DL  LG   A         R +G  V G
baf     3  ILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDLDALGRWLATLPRRPQRALGVNVAG 62

orf61  78  EFKKAQVQEQLAR---KIEWLPSSAQAXGIRNHYRHPEEHGSDRW---FNALGSRRFSRN 131
            +    +    L        I WL +    A G+RN YR+P++ G+DRW         L  +
baf    63  LARGEAIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDQLGADRWACMVGVLARQPSVHP 122

orf61  132 ACVVVSCGTAVTVDALTDDGHYLGXGTIMPGFHLMKESLAVRTANL 177
            +V S GTA T+D +   D   + G G I+PG   +M+ +LA   TA+L
baf    123 PLLVASFGTATTLDTIGPDNVFPG-GLILPGPAMMRGALAYGTAHL 167
```

<u>Homology with a predicted ORF from *N.meningitidis* (strain A)</u>

ORF61 shows 97.4% identity over a 189aa overlap with an ORF (ORF61a) from strain A of *N. meningitidis*:

```
                                                  10        20        30
orf61.pep                                   EISLRSDXRPVSVXKRRDSERFLLLDGGNS
                                            ||||||| ||||| ||||||||||||||||
orf61a       TVFEGTVKGVDGQGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNS
             290       300       310       320       330       340

                     40        50        60        70        80        90
orf61.pep    RLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLAR
             ||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
orf61a       RLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLAR
             350       360       370       380       390       400

                    100       110       120       130       140       150
orf61.pep    KIEWLPSSAQAXGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDD
             ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
orf61a       KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDD
             410       420       430       440       450       460

                    160       170       180     189
orf61.pep    GHYLGXGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPT
             ||||| |||||||||||||||||||||||||||||||||
orf61a       GHYLG-GTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMM
             470       480       490       500       510       520

orf61a       HGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGG
```

```
            530        540        550        560        570        580
```

The complete length ORF61a nucleotide sequence <SEQ ID 235> is:

```
   1 ATGACGGTTT TGAAGCCTTC GCACTGGCGG GTGTTGGCGG AGCTTGCCGA
  51 CGGTTTGCCG CAACACGTCT CGCAACTGGC GCGTATGGCG GATATGAAGC
 101 CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA CATACGCGGG
 151 CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CATTGGCGGT
 201 TTTCGATGCC GAAGGTTTGC GCGAGCTGGG GGAAAGGTCG GGTTTTCAGA
 251 CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
 301 GCGCGGATTG CGCCGGACAA GGCGCACAAA ACCATATGTG TGACCCACCT
 351 GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
 401 GCGAGTGTCT GATGTTCAGT TTTGGCTGGG TGTTTGACCG GCCGCAGTAT
 451 GAGTTGGGTT CGCTGTCGCC TGTTGCGGCA GTGGCGTGCC GGCGCGCCTT
 501 GTCGCGTTTG GGTTTGAAAA CGCAAATCAA GTGGCCAAAC GATTTGGTCG
 551 TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACGGT CAGGACGGGC
 601 GGCAAAACGG TTGCCGTGGT CGGTATCGGC ATCAATTTCG TGCTGCCCAA
 651 GGAAGTGGAA AACGCCGCTT CCGTGCAATC GCTGTTTCAG ACGGCATCGC
 701 GGCGGGGAAA TGCCGATGCC GCCGTGTTGC TGGAAACGCT GTTGGCGGAA
 751 CTTGATGCGG TGTTGTTGCA ATATGCGCGG GACGGATTTG CGCCTTTTGT
 801 GGCGGAATAT CAGGCTGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
 851 TGCGCGACGG CGAAACCGTG TTCGAAGGCA CGGTTAAAGG CGTGGACGGA
 901 CAAGGCGTTC TGCACTTGGA AACGGCAGAG GGCAAACAGA CGGTCGTCAG
 951 CGGCGAAATC AGCCTGCGGT CCGACGACAG GCCGGTTTCC GTGCCGAAGC
1001 GGCGGGATTC GGAACGTTTT CTGCTGTTGG ACGGCGGCAA CAGCCGGCTC
1051 AAGTGGGCGT GGGTGGAAAA CGGCACGTTC GCAACCGTCG GTAGCGCGCC
1101 GTACCGCGAT TTGTCGCCTT TGGGCGCGGA GTGGGCGGAA AAGGTGGATG
1151 GAAATGTCCG CATCGTCGGT TGCGCCGTGT GCGGAGAATT CAAAAAGGCA
1201 CAAGTGCAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC
1251 ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT
1301 CCGACCGCTG GTTCAACGCC TTGGGCAGCC GCCGCTTCAG CCGCAACGCC
1351 TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
1401 TGACGGACAT TATCTCGGGG GAACCATCAT GCCCGGTTTC CACCTGATGA
1451 AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGGCA CGCCGGTAAG
1501 CGTTATCCTT TCCCGACCAC AACGGGCAAT GCCGTCGCCA GCGGCATGAT
1551 GGATGCGGTT TGCGGCTCGG TTATGATGAT GCACGGGCGT TTGAAAGAAA
1601 AAACCGGGGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG
1651 GCAAAAGTTG CCGAAGCCCT GCCGCCTGCA TTTTTGGCGG AAAATACCGT
1701 GCGCGTGGCG GACAACCTCG TCATTCACGG GCTGCTGAAC CTGATTGCCG
1751 CCGAAGGCGG GGAATCGGAA CATACTTAA
```

This encodes a protein having amino acid sequence <SEQ ID 236>:

```
  1 MTVLKPSHWR VLAELADGLP QHVSQLARMA DMKPQQLNGF WQQMPAHIRG
 51 LLRQHDGYWR LVRPLAVFDA EGLRELGERS GFQTALKHEC ASSNDEILEL
101 ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWVFDRPQY
151 ELGSLSPVAA VACRRALSRL GLKTQIKWPN DLVVGRDKLG GILIETVRTG
201 GKTVAVVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLAE
251 LDAVLLQYAR DGFAPFVAEY QAANRDHGKA VLLLRDGETV FEGTVKGVDG
301 QGVLHLETAE GKQTVVSGEI SLRSDDRPVS VPKRRDSERF LLLDGGNSRL
351 KWAWVENGTF ATVGSAPYRD LSPLGAEWAE KVDGNVRIVG CAVCGEFKKA
401 QVQEQLARKI EWLPSSAQAL GIRNHYRHPE EHGSDRWFNA LGSRRFSRNA
451 CVVVSCGTAV TVDALTDDGH YLGGTIMPGF HLMKESLAVR TANLNRHAGK
501 RYPFPTTTGN AVASGMMDAV CGSVMMMHGR LKEKTGAGKP VDVIITGGGA
551 AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAAEGGESE HT*
```

ORF61a and ORF61-1 show 98.5% identity in 591 aa overlap:

```
                   10        20        30        40        50        60
orf61a.pep  MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR
            ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf61-1     MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR
                   10        20        30        40        50        60

                   70        80        90       100       110       120
orf61a.pep  LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf61-1     LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK
                   70        80        90       100       110       120

                  130       140       150       160       170       180
```

```
orf61a.pep   GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN
             ||||||||||||||||||||||||||||||||||||||||||||||||||| :|||||||
orf61-1      GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN
                     130       140       150       160       170       180

                     190       200       210       220       230       240
orf61a.pep   DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf61-1      DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA
                     190       200       210       220       230       240

                     250       260       270       280       290       300
orf61a.pep   AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG
             ||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||||
orf61-1      AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG
                     250       260       270       280       290       300

                     310       320       330       340       350       360
orf61a.pep   QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf61-1      QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF
                     310       320       330       340       350       360

                     370       380       390       400       410       420
orf61a.pep   ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL
             |||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
orf61-1      ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL
                     370       380       390       400       410       420

                     430       440       450       460       470       480
orf61a.pep   GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf61-1      GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF
                     430       440       450       460       470       480

                     490       500       510       520       530       540
orf61a.pep   HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf61-1      HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP
                     490       500       510       520       530       540

                     550       560       570       580       590
orf61a.pep   VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHTX
             |||||||||||||||||||||||||||||||||||:||||:||||| | ||
orf61-1      VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEHIX
                     550       560       570       580       590
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF61 shows 94.2% identity over a 189aa overlap with a predicted ORF (ORF61.ng) from *N. gonorrhoeae:*

```
orf61.pep                                  EISLRSDXRPVSVXKRRDSERFLLLDGGNS    30
                                           ||||| | | ||| || ||||||||:||||
orf61ng      TVCEGTVKGVDGRGVLHLETAEGEQTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNS   211

orf61.pep    RLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLAR    90
             |||||||||||||||||||||||||||||||||||||||||||||||| |||||:|||||
orf61ng      RLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLAR   271

orf61.pep    KIEWLPSSAQAXGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDD   150
             ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
orf61ng      KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDD   331

orf61.pep    GHYLGXGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPT                        189
             ||||| ||||||||||||||||||||||| |||||||||
orf61ng      GHYLG-GTIMPGFHLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMM   390
```

An ORF61ng nucleotide sequence <SEQ ID 237> was predicted to encode a protein having amino acid sequence <SEQ ID 238>:

```
  1 MFSFGWAFDR PQYELGSLSP VAALACRRAL GCLGLETQIK WPNDLVVGRD
 51 KLGGILIETV RAGGKTVAVV GIGINFVLPK EVENAASVQS LFQTASRRGN
101 ADAAVLLETL LAELGAVLEQ YAEEGFAPFL NEYETANRDH GKAVLLLRDG
151 ETVCEGTVKG VDGRGVLHLE TAEGEQTVVS GEISLRPDNR SVSVPKRPDS
201 ERFLLLEGGN SRLKWAWVEN GTFATVGSAP YRDLSPLGAE WAEKADGNVR
251 IVGCAVCGES KKAQVKEQLA RKIEWLPSSA QALGIRNHYR HPEEHGSDRW
301 FNALGSRRFS RNACVVVSCG TAVTVDALTD DGHYLGGTIM PGFHLMKESL
351 AVRTANLNRP AGKRYPFPTT TGNAVASGMM DAVCGSIMMM HGRLKEKNGA
401 GKPVDVIITG GGAAKVAEAL PPAFLAENTV RVADNLVIHG LLNLIAAEGG
451 ESEHA*
```

Further analysis revealed the complete gonococcal DNA sequence <SEQ ID 239> to be:

```
   1 ATGACGGTTT TGAAGCCTTC GCATTGGCGG GTGTTGGCGG AGCTTGCCGA
  51 CGGTTTGCCG CAACACGTAT CGCAATTGGC GCGTGAGGCG GACATGAAGC
 101 CGCAGCAGCT CAACGGTTTT TGGCAGCAGA TGCCGGCGCA TATACGCGGG
 151 CTGTTGCGCC AACACGACGG CTATTGGCGG CTGGTGCGCC CCTTGGCGGT
 201 TTTCGATGCC GAAGGTTTGC GCGATCTGGG GGAAAGGTCG GGTTTTCAGA
 251 CGGCATTGAA GCACGAGTGC GCGTCCAGCA ACGACGAGAT ACTGGAATTG
 301 GCGCGGATTG CGCCGGACAA GGCGCACAAA ACCATATGCG TGACCCACCT
 351 GCAAAGTAAG GGCAGGGGGC GGCAGGGGCG GAAGTGGTCG CACCGTTTGG
 401 GCGAGTGCCT GATGTTCAGT TTCGGCTGGG CGTTTGACCG GCCGCAGTAT
 451 GAGTTGGGTT CGCTGTCGCC TGTTGCGGCA CTTGCGTGCC GGCGCGCTTT
 501 GGGGTGTTTG GGTTTGGAAA CGCAAATCAA GTGGCCAAAC GATTTGGTCG
 551 TCGGACGCGA CAAATTGGGC GGCATTCTGA TTGAAACAGT CAGGGCGGGC
 601 GGTAAAACGG TTGCCGTGGT CGGTATCGGC ATCAATTTCG TGCTGCCCAA
 651 GGAAGTGGAA AACGCCGCTT CCGTGCAGTC GCTGTTTCAG ACGGCATCGC
 701 GGCGGGGCAA TGCCGATGCC GCCGTATTGC TGGAAACATT GCTTGCGGAA
 751 CTGGGCGCGG TGTTGGAACA ATATGCGGAA GAAGGGTTCG CGCCATTTTT
 801 AAATGAGTAT GAAACGGCCA ACCGCGACCA CGGCAAGGCG GTATTGCTGT
 851 TGCGCGACGG CGAAACCGTG TGCGAAGGCA CGGTTAAAGG CGTGGACGGA
 901 CGAGGCGTTC TGCACTTGGA AACGGCAgaa ggcgaACAGa cggtcgtcag
 951 cggcgaaaTC AGcctGCggc ccgacaacaG GTCGGtttcc gtgccgaagc
1001 ggccggatTC GgaacgtTTT tTGCtgttgg aaggcgggaa cagccgGCTC
1051 AAGTGGGCGT GggtggAAAa cggcacgttc gcaaccgtgg gcagcgcgCc
1101 gtaCCGCGAT TTGTCGCCTT TGGGCGCGGA GTGGGCGGAA AAGGCGGATG
1151 GAAATGTCCG CATCGTCGGT TGCGCCGTGT GCGGAGAATC CAAAAAGGCA
1201 CAAGTGAAGG AACAGCTCGC CCGAAAAATC GAGTGGCTGC CGTCTTCCGC
1251 ACAGGCTTTG GGCATACGCA ACCACTACCG CCACCCCGAA GAACACGGTT
1301 CCGACCGTTG GTTCAACGCC TTGGGCAGCC GCCGCTTCAG CCGCAACGCC
1351 TGCGTCGTCG TCAGTTGCGG CACGGCGGTA ACGGTTGACG CGCTCACCGA
1401 TGACGGACAT TATCTCGGCG GAACCATCAT GCCCGGCTTC CACCTGATGA
1451 AAGAATCGCT CGCCGTCCGA ACCGCCAACC TCAACCGCCC CGCCGGCAAA
1501 CGTTACCCTT TCCCGACCAC AACGGGCAAC GCCGTCGCAA GCGGCATGAT
1551 GGACGCGGTT TGCGGCTCGA TAATGATGAT GCACGGCCGT TTGAAAGAAA
1601 AAAACGGCGC GGGCAAGCCT GTCGATGTCA TCATTACCGG CGGCGGCGCG
1651 GCGAAAGTCG CCGAAGCCCT GCCGCCTGCA TTTTTGGCGG AAAATACCGT
1701 GCGCGTGGCG GACAACCTCG TCATCCACGG GCTGCTGAAC CTGATTGCCG
1751 CCGAAGGCGG GGAATCGGAA CACGCTTAA
```

This corresponds to the amino acid sequence <SEQ ID 240; ORF61ng-1>:

```
  1 MTVLKPSHWR VLAELADGLP QHVSQLAREA DMKPQQLNGF WQQMPAHIRG
 51 LLRQHDGYWR LVRPLAVFDA EGLRDLGERS GFQTALKHEC ASSNDEILEL
101 ARIAPDKAHK TICVTHLQSK GRGRQGRKWS HRLGECLMFS FGWAFDRPQY
151 ELGSLSPVAA LACRRALGCL GLETQIKWPN DLVVGRDKLG GILIETVRAG
201 GKTVAVVGIG INFVLPKEVE NAASVQSLFQ TASRRGNADA AVLLETLLAE
251 LGAVLEQYAE EGFAPFLNEY ETANRDHGKA VLLLRDGETV CEGTVKGVDG
301 RGVLHLETAE GEQTVVSGEI SLRPDNRSVS VPKRPDSERF LLLEGGNSRL
351 KWAWVENGTF ATVGSAPYRD LSPLGAEWAE KADGNVRIVG CAVCGESKKA
401 QVKEQLARKI EWLPSSAQAL GIRNHYRHPE EHGSDRWFNA LGSRRFSRNA
451 CVVVSCGTAV TVDALTDDGH YLGGTIMPGF HLMKESLAVR TANLNRPAGK
501 RYPFPTTTGN AVASGMMDAV CGSIMMMHGR LKEKNGAGKP VDVIITGGGA
551 AKVAEALPPA FLAENTVRVA DNLVIHGLLN LIAAEGGESE HA*
```

ORF61ng-1 and ORF61-1 show 93.9% identity in 591 aa overlap:

```
orf61ng-1.pep  MTVLKPSHWRVLAELADGLPQHVSQLAREADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR  60
               ||||| |||||||||||||||||||||| |||||||||||||||||||||||||||||||
orf61-1        MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR  60

orf61ng-1.pep  LVRPLAVFDAEGLRDLGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
               ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
orf61-1        LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

orf61ng-1.pep  GRGRQGRKWSHRLGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPN 180
               |||||||||||||||||||||||:||||||||||||||||:||||||: |||::||||||
orf61-1        GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180

orf61ng-1.pep  DLVVGRDKLGGILIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
               ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
orf61-1        DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

orf61ng-1.pep  AVLLETLLAELGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDG 300
               ||||||||:|| ||| |||::|||||: ||::|||||||||||||||||| |||||||||
orf61-1        AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

orf61ng-1.pep  RGVLHLETAEGEQTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTF 360
               :||||||||||:||||||||||| |:| |||||| ||||||||:||||||||||||||||
orf61-1        QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

orf61ng-1.pep  ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQAL 420
               |||||||||||||||||||||||||||||||||||| |||||:|||||||||||||||||
orf61-1        ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

orf61ng-1.pep  GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf61-1        GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

orf61ng-1.pep  HLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKP 540
               |||||||||||||||| ||||||||||||||||||||||||||:||||||||||:|||||
orf61-1        HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

orf61ng-1.pep  VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHAX   593
               ||||||||||||||||||||||||||||||||||| ||||:||||| | ||
orf61-1        VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEHIX   593
```

Based on this analysis, including the homology with the baf protein of *B.pertussis* and the presence of a putative prokaryotic membrane lipoprotein lipid attachment site, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 29

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 241>:

```
  1  ATGTTTTACC AAATCCTTGC CCTGATTATC TGGAGCAGCT CGTTTATTGC
 51  CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG GTCGGCGTGC
101  GCCTGCTAAT TGCCGCGCTG CCTGCACTGC CCGCCTGCCG CCGTCATGTC
151  GGCAAGATTC CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTCGT
201  CAACTATGTG CTGACCCTGC TGCTTCAGTT TGTCGGGTTG AAATACACTT
251  CCGCCGCCAG CGCATCGGTC ATTGTCGGAC TCGAGCCGCT GCTGATGGTG
301  TTTGTCGGAC ACTTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT
351  ATGCGGCGCG GCGGCATTTG CCGGTGTCGC GCTGCTGATG GCGGGCGGTG
401  CGGaAGAGGG CGGCGaAGTC GGCTGGTTCG GCTGCCTGCT GGTGTTGTTG
451  GCGGGCGCGG GCTTTTGTGC CGCTATGCGT CCGACGCAAA GGCTGATTGC
501  ACGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT
551  TGATGTGCCT GCCGTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGGAC
601  TGGAGCGTCG GGATGGTATT GTCGCTGCTG TATTTGGGTT TGGGGTGC..
```

This corresponds to the amino acid sequence <SEQ ID 242; ORF62>:

```
  1  MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLLIAAL PALPACRRHV
 51  GKIPREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV
101  FVGHFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL
151  AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
201  WSVGMVLSLL YLGLGC..
```

Further work revealed the complete nucleotide sequence <SEQ ID 243>:

```
  1  ATGTTTTACC AAATCCTTGC CCTGATTATC TGGAGCAGCT CGTTTATTGC
 51  CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG GTCGGCGTGC
101  GCCTGCTAAT TGCCGCGCTG CCTGCACTGC CCGCCTGCCG CCGTCATGTC
151  GGCAAGATTC CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTCGT
201  CAACTATGTG CTGACCCTGC TGCTTCAGTT TGTCGGGTTG AAATACACTT
251  CCGCCGCCAG CGCATCGGTC ATTGTCGGAC TCGAGCCGCT GCTGATGGTG
301  TTTGTCGGAC ACTTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT
351  ATGCGGCGCG GCGGCATTTG CCGGTGTCGC GCTGCTGATG GCGGGCGGTG
401  CGGAAGAGGG CGGCGAAGTC GGCTGGTTCG GCTGCCTGCT GGTGTTGTTG
451  GCGGGCGCGG GCTTTTGTGC CGCTATGCGT CCGACGCAAA GGCTGATTGC
501  ACGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT
551  TGATGTGCCT GCCGTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGGAC
601  TGGAGCGTCG GGATGGTATT GTCGCTGCTG TATTTGGGTT TGGGGTGCGG
651  CTGGTACGCC TATTGGCTGT GGAACAAGGG GATGAGCCGT GTTCCTGCCA
701  ATGTTTCGGG ACTGTTGATT TCGCTCGAAC CCGTCGTCGG CGTGCTGCTG
751  GCGGTTTTGA TTTTGGGCGA ACACCTGTCG CCCGTGTCCG CCTTGGGCGT
801  GTTTGTCGTC ATCGCCGCCA CCTTGGTTGC CGGCCGGCTG TCGCATCAAA
851  AATAA
```

This corresponds to the amino acid sequence <SEQ ID 244; ORF62-1>:

```
  1  MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLLIAAL PALPACRRHV
 51  GKIPREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV
101  FVGHFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL
151  AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
201  WSVGMVLSLL YLGLGCGWYA YWLWNKGMSR VPANVSGLLI SLEPVVGVLL
251  AVLILGEHLS PVSALGVFVV IAATLVAGRL SHQK*
```

Computer analysis of this amino acid sequence gave the following results:

<u>Homology with hypothetical transmembrane protein HI0976 of *H. influenzae* (accession number Q57147)</u>

ORF62 and HI0976 show 50% aa identity in 114aa overlap:

```
Orf62    1   MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRXXXXXXXXXXXCRRHVGKIPREEWKP 60
             M YQILAL+IWSSS I  K  Y  +DP L+V VR           R    KI +    K
HI0976   1   MLYQILALLIWSSSLIVGKLTYSMMDPVLVVQVRLIIAMIIVMPLFLRRWKKIDKPMRKQ 60

Orf62   61   LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAY 114
             L  ++F NY    LLQF+GLKYTSA+SA  ++GLEPLL+VFVGHFFF  K   +
HI0976  61   LWWLAFFNYTAVFLLQFIGLKYTSASSAVTMIGLEPLLVVFVGHFFFKTKQNGF 114
```

<u>Homology with a predicted ORF from *N.meningitidis* (strain A)</u>

ORF62 shows 99.5% identity over a 216aa overlap with an ORF (ORF62a) from strain A of *N. meningitidis:*

```
                   10        20        30        40        50        60
orf62.pep   MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf62a      MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP
                   10        20        30        40        50        60

                   70        80        90       100       110       120
orf62.pep   LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAYHWICGA
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf62a      LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAYHWICGA
                   70        80        90       100       110       120

                  130       140       150       160       170       180
orf62.pep   AAFAGVALLMAGGAEEGGEVGWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf62a         AAFAGVALLMAGGAEEGGEVGWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA
                   130       140       150       160       170       180

                   190       200       210
orf62.pep      AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGC
               ||||||||||||||||||||||||||||||||||:||
orf62a         AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVGCSWYAYWLWNKGMSRVPANVSGLLI
                   190       200       210       220       230       240

orf62a         SLEPVVGVLLAVLILGEHLSPVSVLGVFVVIAATLVAGRLSHQKX
                   250       260       270       280
```

The complete length ORF62a nucleotide sequence <SEQ ID 245> is:

```
  1  ATGTTTTACC AAATCCTTGC CCTGATTATC TGGAGCAGCT CGTTTATTGC
 51  CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG GTCGGCGTGC
101  GCCTGCTGAT TGCTGCGCTG CCTGCACTGC CCGCCTGCCG CCGTCATGTC
151  GGCAAGATTC CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTCGT
201  CAACTATGTG CTGACCCTGC TACTTCAGTT TGTCGGGTTG AAATACACTT
251  CCGCCGCCAG CGCATCGGTC ATTGTCGGAC TCGAGCCACT GCTGATGGTG
301  TTTGTCGGAC ACTTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT
351  ATGCGGCGCG GCGGCATTTG CCGGTGTCGC GCTGCTGATG GCGGGCGGTG
401  CGGAAGAGGG CGGCGAAGTC GGCTGGTTCG GCTGCCTGCT GGTGTTGTTG
451  GCGGGCGCGG GCTTTTGTGC CGCTATGCGT CCGACGCAAA GGCTGATTGC
501  ACGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT
551  TGATGTGCCT GCCGTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGGAC
601  TGGAGCGTCG GAATGGTATT GTCGCTGCTG TATTTGGGCG TGGGGTGCAG
651  CTGGTACGCC TATTGGCTGT GGAACAAGGG GATGAGCCGT GTTCCTGCCA
701  ACGTTTCGGG ACTGTTGATT TCGCTCGAAC CCGTCGTCGG CGTGCTGCTG
751  GCGGTTTTGA TTTTGGGCGA ACACCTGTCG CCCGTGTCCG TCTTGGGCGT
801  GTTTGTCGTC ATCGCCGCCA CCTTGGTTGC CGGCCGGCTG TCGCATCAAA
851  AATAA
```

This encodes a protein having amino acid sequence <SEQ ID 246>:

```
  1  MFYQILALII WSSSFIAAKY VYGGIDPALM VGVRLLIAAL PALPACRRHV
 51  GKIPREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV
101  FVGHFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL
151  AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
201  WSVGMVLSLL YLGVGCSWYA YWLWNKGMSR VPANVSGLLI SLEPVVGVLL
251  AVLILGEHLS PVSVLGVFVV IAATLVAGRL SHQK*
```

ORF62a and ORF62-1 show 98.9% identity in 284 aa overlap:

```
orf62a.pep   MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf62-1      MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP  60

orf62a.pep   LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAYHWICGA 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf62-1      LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAYHWICGA 120

orf62a.pep   AAFAGVALLMAGGAEEGGEVGWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf62-1      AAFAGVALLMAGGAEEGGEVGWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA 180

orf62a.pep   AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGVGCSWYAYWLWNKGMSRVPANVSGLLI 240
             ||||||||||||||||||||||||||||||||||:||:||||||||||||||||||||||
orf62-1      AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANVSGLLI 240

orf62a.pep   SLEPVVGVLLAVLILGEHLSPVSVLGVFVVIAATLVAGRLSHQKX  285
             ||||||||||||||||||||||||:||||||||||||||||||||
orf62-1      SLEPVVGVLLAVLILGEHLSPVSALGVFVVIAATLVAGRLSHQKX  285
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF62 shows 99.5% identity over a 216aa overlap with a predicted ORF (ORF62.ng) from *N. gonorrhoeae:*

```
orf62.pep   MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP   60
            ||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||
orf62ng     MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP   60

orf62.pep   LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAYHWICGA  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf62ng     LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAYHWICGA  120

orf62.pep   AAFAGVALLMAGGAEEGGEVGWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf62ng     AAFAGVALLMAGGAEEGGEVGWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA  180

orf62.pep   AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGC                          216
            ||||||||||||||||||||||||||||||||||||
orf62ng     AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI  240
```

The complete length ORF62ng nucleotide sequence <SEQ ID 247> is:

```
  1 ATGTTTTACC AAATCCTTGC CCTGATTATC TGGGGCAGCT CGTTTATTGC
 51 CGCCAAATAT GTCTATGGCG GCATCGATCC CGCATTGATG GTCGGCGTGC
101 GCCTGCTGAT TGCCGCGCTG CCTGCACTGC CCGCCTGCCG CCGTCATGTC
151 GGCAAGATTC CGCGTGAGGA ATGGAAGCCG TTGCTGATTG TGTCGTTCGT
201 CAACTATGTG CTGACCCTGC TGCTTCAGTT TGTCGGGTTG AAATACACTT
251 CCGCCGCCAG CGCATCGGTC ATTGTCGGAC TCGAGCCGCT GCTGATGGTG
301 TTTGTCGGAC ACTTTTTCTT CAACGACAAA GCGCGTGCCT ACCACTGGAT
351 ATGCGGCGCG GCGGCATTTG CCGGTGTCGC GCTGCTGATG GCGGGCGGTG
401 CGGAAGAGGG CGGCGAAGTC GGCTGGTTCG GCTGCCTGCT GGTGTTGTTG
451 GCGGGCGCGG GCTTTTGTGC CGCTATGCGT CCGACGCAAA GGCTGATTGC
501 CCGCATCGGC GCACCGGCAT TCACATCTGT TTCCATTGCC GCCGCATCGT
551 TGATGTGCCT GCCGTTTTCG CTTGCTTTGG CGCAAAGTTA TACCGTGGAC
601 TGGAGCGTCG GGATGGTATT GTCGCTGTTG TATTTGGGTT TGGGGTGCGG
651 CTGGTACGCC TATTGGCTGT GGAACAAGGG GATGAGCCGT GTTCCTGCCA
701 ACGCGTCGGG ACTGTTGATT TCGCTCGAAC CCGTCGTCGG CGTGCTGTTG
751 GCGGTTTTGA TTTTGGGCGA ACATTTATCG CCCGTGTCCG CCTTGGGCGT
801 GTTTGTCGTC ATCGCCGCCA CTTTCGCCGC CGGCCGGCTG TCGCGCAGGG
851 ACGCGCAAAA CGGCAATGCC GTCTGA
```

This encodes a protein having amino acid sequence <SEQ ID 248>:

```
  1 MFYQILALII WGSSFIAAKY VYGGIDPALM VGVRLLIAAL PALPACRRHV
 51 GKIPREEWKP LLIVSFVNYV LTLLLQFVGL KYTSAASASV IVGLEPLLMV
101 FVGHFFFNDK ARAYHWICGA AAFAGVALLM AGGAEEGGEV GWFGCLLVLL
151 AGAGFCAAMR PTQRLIARIG APAFTSVSIA AASLMCLPFS LALAQSYTVD
201 WSVGMVLSLL YLGLGCGWYA YWLWNKGMSR VPANASGLLI SLEPVVGVLL
251 AVLILGEHLS PVSALGVFVV IAATFAAGRL SRRDAQNGNA V*
```

ORF62ng and ORF62-1 show 97.9% identity in 283 aa overlap:

```
                    10        20        30        40        50        60
orf62ng.pep  MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP
             ||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||
orf62-1      MFYQILALIIWSSSFIAAKYVYGGIDPALMVGVRLLIAALPALPACRRHVGKIPREEWKP
                    10        20        30        40        50        60

                    70        80        90       100       110       120
orf62ng.pep  LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAYHWICGA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf62-1      LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAYHWICGA
                    70        80        90       100       110       120

                   130       140       150       160       170       180
orf62ng.pep  AAFAGVALLMAGGAEEGGEVGWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf62-1      AAFAGVALLMAGGAEEGGEVGWFGCLLVLLAGAGFCAAMRPTQRLIARIGAPAFTSVSIA
                   130       140       150       160       170       180

                   190       200       210       220       230       240
orf62ng.pep  AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANASGLLI
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||:||||
orf62-1      AASLMCLPFSLALAQSYTVDWSVGMVLSLLYLGLGCGWYAYWLWNKGMSRVPANVSGLLI
                   190       200       210       220       230       240
```

```
                     250       260       270       280       290
orf62ng.pep    SLEPVVGVLLAVLILGEHLSPVSALGVFVVIAATFAAGRLSRRDAQNGNAVX
               ||||||||||||||||||||||||||||||||||::|||||::
orf62-1        SLEPVVGVLLAVLILGEHLSPVSALGVFVVIAATLVAGRLSHQKX
                     250       260       270       280
```

Furthermore, ORF62ng shows significant homology to a hypothetical *H.influenzae* protein:

```
sp|Q57147|Y976_HAEIN HYPOTHETICAL PROTEIN HI0976 >gi|1074589|pir||B64163
hypothetical protein HI0976 - Haemophilus influenzae (strain Rd KW20)
>gi|1574004 (U32778) hypothetical [Haemophilus influenzae] Length = 128
 Score =  106 bits (262), Expect = 2e-22
 Identities = 56/114 (49%), Positives = 68/114 (59%)

Query: 1   MFYQILALIIWGSSFIAAKYVYGGIDPALMVGVRXXXXXXXXXXXCRRHVGKIPREEWKP 60
            M YQILAL+IW SS I  K  Y  +DP L+V VR            R   KI +   K
Sbjct: 1   MLYQILALLIWSSSLIVGKLTYSMMDPVLVVQVRLIIAMIIVMPLFLRRWKKIDKPMRKQ 60

Query: 61  LLIVSFVNYVLTLLLQFVGLKYTSAASASVIVGLEPLLMVFVGHFFFNDKARAY 114
            L  ++F NY    LLQF+GLKYTSA+SA  ++GLEPLL+VFVGHFFF  K   +
Sbjct: 61  LWWLAFFNYTAVFLLQFIGLKYTSASSAVTMIGLEPLLVVFVGHFFFKTKQNGF 114
```

Based on this analysis, including the homology with the transmembrane protein of *H.influenzae* and the putative leader sequecne and several transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 30

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 249>:

```
   1  ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCmGwms TCCTGkkGTA
  51  sGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
 101  GGTGGATTGT TGCGTTCAGC GCAATGCTGC TGCTGGTGTT GTCCGCCGTT
 151  TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT
 201  CGGTTCGCtA srTyGCCAAA gsGCCTgkks TGGG.ATGTT TACGCTGGTT
 251  GCCGkACTGC CCGGCGTGTT TCTGTTCGGC TTTCCCGCAC AGTTCATCAA
 301  CGGCACGATT AATTCGTGGT TCGGCAACGA TACCCACGAG GCGCTTGAAC
 351  GCAGCCTCAA TTTGAGCAAG TCCGCATTGA ATTTGGCGGC AGACAACGCC
 401  CTCGGCAACG CCGTCCCCGT GCAGATAGAC CTCATCGGCG CGGCTTCCCT
 451  GCCCGGGGAT ATGGGCAGGG TGCTGGAACA TTACGCCGGC AGCGGTTTTG
 501  CCCAGCTTGC CCTGTACAAy ksCGCAAGCG GCAAAATCGA AAAAAGCATC
 551  AACCCGCACA AGCTCGATCA GCCGTTTCCA GGTAAGGCGC GTTGGGAaAa
 601  AATCCaACGG GCGGGTTCGG TCAGGGATTT GGAAAGCATA GGCGGCGTAT
 651  TGTaCGCGCA GGGCTGGCTG TCGGCGGGTA CGCACwACGG GCGCGATTAC
 701  GCCTTGTTTT TCCGTCAGCC GGTTCCCAAA GGCGTGGCAG AGGATGCCGT
 751  yTTAATCGAA AAGGCAAGGG CGAAATATGC TGAGTTGAGT TACAGCAAAA
 801  AAGGTTTGCA GACCTTTTTC CTGGCAACCC TGCTGATTGC CTCGCTGCTG
 851  TCGATTTTTC TTGCACTGGT CATGGCACTG TATTTCGCCC GCCGTTTCGT
 901  CGAACCCGTC CTATCGCTTG CCGAGGGGGC GAAGGCGGTG GCGCAAGGCG
 951  ATTTCAGCCA GACGCGCCCC GTGTTGCGCA ACGACGAGTT CGGACGCTTG
1001  ACCArGTTGT TCAACCACAT GACCGAGCAG CTTTCCATCG CCAAAGATGC
1051  AGACGAGCGC AACCGCCGGC GCGAGGAAGC CGCCAGGCAT TATCTTGAAT
1101  GCGTGTTGGA GGGGCTGACC ACGGGCGTGG TGGTGTTTGA CGAACAAGGC
1151  TGTCTGAAAA CCTTCAACAA AGCGGCGGGT ACC..
```

This corresponds to the amino acid sequence <SEQ ID 250; ORF64>:

```
   1  MRRFLPIAAI CAXXLXXGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
  51  LARYVILLLK DRRDGVFGSX XAKXPXXXMF TLVAXLPGVF LFGFPAQFIN
 101  GTINSWFGND THEALERSLN LSKSALNLAA DNALGNAVPV QIDLIGAASL
 151  PGDMGRVLEH YAGSGFAQLA LYNXASGKIE KSINPHKLDQ PFPGKARWEK
 201  IQRAGSVRDL ESIGGVLYAQ GWLSAGTHXG RDYALFFRQP VPKGVAEDAV
 251  LIEKARAKYA ELSYSKKGLQ TFFLATLLIA SLLSIFLALV MALYFARRFV
```

```
301  EPVLSLAEGA KAVAQGDFSQ TRPVLRNDEF GRLTXLFNHM TEQLSIAKDA
351  DERNRRREEA ARHYLECVLE GLTTGVVVFD EQGCLKTFNK AAGT..
```

Further work revealed the complete nucleotide sequence <SEQ ID 251>:

```
   1  ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCCGTCG TCCTGTTGTA
  51  CGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
 101  GGTGGATTGT TGCGTTCAGC GCAATGCTGC TGCTGGTGTT GTCCGCCGTT
 151  TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT
 201  CGGTTCGCAG ATTGCCAAAC GCCTTTCTGG GATGTTTACG CTGGTTGCCG
 251  TACTGCCCGG CGTGTTTCTG TTCGGCGTTT CCGCACAGTT CATCAACGGC
 301  ACGATTAATT CGTGGTTCGG CAACGATACC CACGAGGCGC TTGAACGCAG
 351  CCTCAATTTG AGCAAGTCCG CATTGAATTT GGCGGCAGAC AACGCCCTCG
 401  GCAACGCCGT CCCCGTGCAG ATAGACCTCA TCGGCGCGGC TTCCCTGCCC
 451  GGGGATATGG GCAGGGTGCT GGAACATTAC GCCGGCAGCG GTTTTGCCCA
 501  GCTTGCCCTG TACAATGCCG CAAGCGGCAA AATCGAAAAA AGCATCAACC
 551  CGCACAAGCT CGATCAGCCG TTTCCAGGTA AGGCGCGTTG GGAAAAAATC
 601  CAACGGGCGG GTTCGGTCAG GGATTTGGAA AGCATAGGCG GCGTATTGTA
 651  CGCGCAGGGC TGGCTGTCGG CGGGTACGCA CAACGGGCGC GATTACGCCT
 701  TGTTTTTCCG TCAGCCGGTT CCCAAAGGCG TGGCAGAGGA TGCCGTCTTA
 751  ATCGAAAAGG CAAGGGCGAA ATATGCTGAG TTGAGTTACA GCAAAAAAGG
 801  TTTGCAGACC TTTTTCCTGG CAACCCTGCT GATTGCCTCG CTGCTGTCGA
 851  TTTTTCTTGC ACTGGTCATG GCACTGTATT TCGCCCGCCG TTTCGTCGAA
 901  CCCGTCCTAT CGCTTGCCGA GGGGGCGAAG GCGGTGGCGC AAGGCGATTT
 951  CAGCCAGACG CGCCCCGTGT TGCGCAACGA CGAGTTCGGA CGCTTGACCA
1001  AGTTGTTCAA CCACATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC
1051  GAGCGCAACC GCCGGCGCGA GGAAGCCGCC AGGCATTATC TTGAATGCGT
1101  GTTGGAGGGG CTGACCACGG GCGTGGTGGT GTTTGACGAA CAAGGCTGTC
1151  TGAAAACCTT CAACAAAGCG GCGGAACAGA TTTTGGGGAT GCCGCTTACC
1201  CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA
1251  GTCCCTGCTT GCCGAAGTGT TTGCCGCCAT CGGCGCGGCG GCAGGTACGG
1301  ACAAACCGGT CCATGTGAAA TATGCCGCGC CGGACGATGC CAAAATCCTG
1351  CTGGGCAAGG CAACCGTCCT GCCCGAAGAC AACGGCAACG GCGTGGTAAT
1401  GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAAAAA GAAGCCGCGT
1451  GGGGCGAAGT GGCGAAGCGG CTGGCACACG AAATCCGCAA TCCGCTCACG
1501  CCCATCCAGC TTTCCGCCGA ACGGCTGGCG TGGAAATTGG GCGGGAAGCT
1551  GGATGAGCAG GATGCGCAAA TCCTGACGCG TTCGACCGAC ACCATCGTCA
1601  AACAGGTGGC GGCATTGAAG GAAATGGTCG AAGCATTCCG CAATTATGCG
1651  CGTTCCCCTT CGCTCAAATT GGAAAATCAG GATTTGAACG CCTTAATCGG
1701  CGATGTGTTG GCATTGTATG AAGCCGGTCC GTGCCGGTTT GCGGCGGAGC
1751  TTGCCGGCGA ACCGCTGACG GTGGCGGCGG ATACGACCGC CATGCGGCAG
1801  GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
1851  TGTGCCCGAA GTCAGGGTAA AATCGGAAAC AGGGCAGGAC GGTCGGATTG
1901  TCCTGAGGGT TTGGGACAAC GGCAAAGGGT TCGGCAGGGA AATGCTGCAC
1951  AACGCCTTGG AGCCGTATGT AACGGACAAA CCGGCGGGAA CGGGATTGGG
2001  TCTGCCTGTG GTGAAAAAAA TCATTGAAGA ACACGGGGGC CGCATCAGCC
2051  TGAGCAATCA GGATGCGGGT GGCGGGTGTG TCAGAATGAT CTTGGCAAAA
2101  ACGGTAAAAA CTTATGCGTA G
```

This corresponds to the amino acid sequence <SEQ ID 252; ORF64-1>:

```
  1  MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
 51  LARYVILLLK DRRDGVFGSQ IAKRLSGMFT LVAVLPGVFL FGVSAQFING
101  TINSWFGNDT HEALERSLNL SKSALNLAAD NALGNAVPVQ IDLIGAASLP
151  GDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQP FPGKARWEKI
201  QRAGSVRDLE SIGGVLYAQG WLSAGTHNGR DYALFFRQPV PKGVAEDAVL
251  IEKARAKYAE LSYSKKGLQT FFLATLLIAS LLSIFLALVM ALYFARRFVE
301  PVLSLAEGAK AVAQGDFSQT RPVLRNDEFG RLTKLFNHMT EQLSIAKEAD
351  ERNRRREEAA RHYLECVLEG LTTGVVVFDE QGCLKTFNKA AEQILGMPLT
401  PLWGSSRHGW HGVSAQQSLL AEVFAAIGAA AGTDKPVHVK YAAPDDAKIL
451  LGKATVLPED NGNGVVMVID DITVLIHAQK EAAWGEVAKR LAHEIRNPLT
501  PIQLSAERLA WKLGGKLDEQ DAQILTRSTD TIVKQVAALK EMVEAFRNYA
551  RSPSLKLENQ DLNALIGDVL ALYEAGPCRF AAELAGEPLT VAADTTAMRQ
601  VLHNIFKNAA EAAEEADVPE VRVKSETGQD GRIVLTVCDN GKGFGREMLH
651  NAFEPYVTDK PAGTGLGLPV VKKIIEEHGG RISLSNQDAG GACVRIILPK
701  TVKTYA*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF64 shows 92.6% identity over a 392aa overlap with an ORF (ORF64a) from strain A of *N. meningitidis:*

```
                        10        20        30        40        50        60
orf64.pep     MRRFLPIAAICAXXLXXGLTAATGSTSSLADYFWWIVAFSAMLLLVLSAVLARYVILLLK
              |||||||||||   |  |||||||||||||||||||||||||||||||||||||||||||
orf64a        MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVAFSAMLLLVLSAVLARYVILLLK
                        10        20        30        40        50        60

                        70        80        90       100       110       120
orf64.pep     DRRDGVFGSXXAKXPXXXMFTLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERSLN
              |||||||||  ||     |||||| |||||||| |||||||||||||||||||||||||
orf64a        DRRDGVFGSQIAKR-LSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLN
                        70        80        90       100       110

                       130       140       150       160       170       180
orf64.pep     LSKSALNLAADNALGNAVPVQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNXASGKIE
              ||||||||||||||||:||||| ||||||| |||||||||||||||||||| ||||||
orf64a        LSKSALNLAADNALGNAIPVQIDXIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIE
           120       130       140       150       160       170

                       190       200       210       220       230       240
orf64.pep     KSINPHKLDQPFPGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFFRQP
              |||||||||||||||||||||:|||||| ||||||||| ||||| || |||||||||||
orf64a        KSINPHKLDQPFPGKARWEKIQQAGSVRDXESIGGVLYAXGWLSAXTHNGRDYALFFRQP
           180       190       200       210       220       230

                       250       260       270       280       290       300
orf64.pep     VPKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV
              ||||||||||||||||    ||||||||||||||||||||||||||||||||||||||||
orf64a        VPKGVAEDAVLIEKARAXXXXLSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV
           240       250       260       270       280       290

                       310       320       330       340       350       360
orf64.pep     EPVLSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTXLFNHMTEQLSIAKDADERNRRREEA
              |||||||||||||||||||||||||||||||||| ||||||||||||||:||||||||||
orf64a        EPVLSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTKLFNHMTEQLSIAKEADERNRRREEA
           300       310       320       330       340       350

                       370       380       390
orf64.pep     ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAAGT
              ||||||||||||||||||||||||||||||||
orf64a        ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAAEQILGMPLTPLWGSSRHGWHGVSAQQSL
           360       370       380       390       400       410

orf64a        LAEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNXNGVVMVIDDITVLIHAQ
           420       430       440       450       460       470
```

The complete length ORF64a nucleotide sequence <SEQ ID 253> is:

```
  1 ATGCGCCGTT TTCTACCGAT CGCAGCCATA TGCGCCGTCG TCCTGTTGTA
 51 CGGACTGACG GCGGCAACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
101 GGTGGATTGT TGCGTTCAGC GCAATGCTGC TGCTGGTGTT GTCCGCCGTT
151 TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCG ACGGCGTATT
201 CGGTTCGCAG ATTGCCAAAC GCCTTTCCGG GATGTTTACG CTGGTTGCCG
251 TACTGCCCGG CGTGTTTCTG TTCGGCGTTT CCGCACAGTT TATCAACGGC
301 ACGATTAATT CGTGGTTCGG CAACGATACC CACGAGGCGC TTGAACGCAG
351 CCTCAATTTG AGCAAGTCCG CATTGAATCT GGCGGCAGAC AACGCCCTTG
401 GCAACGCCAT CCCCGTGCAG ATAGACNTCA TCGGCGCGGC TTCCCTGCCC
451 NGGGATATGG GCAGGGTGCT GGAACATTAC GCCGGCAGCG GTTTTGCCCA
501 GCTTGCCCTG TACAATGCCG CAAGCGGCAA AATCGAAAAA AGCATCAACC
551 CGCACAAGCT CGATCAGCCG TTTCCAGGTA AGGCGCGTTG GGAAAAAATC
601 CAACAGGCGG GTTCGGTCAG GGATNNGGAA AGCATAGGCG GCGTATTGTA
651 CGCGCANGGC TGGCTGTCGG CAGNNACGCA CAACGGGCGC GATTACGCCT
701 TGTTTTTCCG TCAGCCGGTT CCCAAAGGCG TGGCAGAGGA TGCCGTCTTA
751 ATCGAAAAGG CAAGGGCGNA ANANNNTNAG TTGAGTTACA GCAAAAAAGG
801 TTTGCAGACC TTTTTCCTNG CAACCCTGCT GATTGCCTCN CTGCTGTCGA
851 TTTTTCTTGC ACTGGTCATG GCACTGTATT TCGCCCGCCG TTTCGTCGAA
```

```
 901  CCCGTCCTAT CGCTTGCCGA GGGGGCGAAG GCGGTGGCGC AAGGCGATTT
 951  CAGCCAGACG CGCCCCGTGT TGCGCAACGA CGAGTTCGGA CGCTTGACCA
1001  AGTTGTTCAA CCACATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC
1051  GAGCGCAACC GCCGGCGCGA GGAAGCCGCC AGACATTATC TCGAATGCGT
1101  GTTGGAGGGG CTGACCACGG GCGTGGTGGT GTTTGACGAA CAAGGCTGTC
1151  TGAAAACCTT CAACAAAGCG GCGGAACAGA TTTTGGGGAT GCCGCTTACC
1201  CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA
1251  GTCCCTGCTT GCCGAAGTGT TTGCCGCCAT CGGCGCGGCG GCAGGTACGG
1301  ACAAACCGGT CCATGTGAAA TATGCCGCGC CGGACGATGC CAAAATCCTG
1351  CTGGGCAAGG CAACCGTCCT GCCCGAAGAC AACNGCAACG GCGTGGTAAT
1401  GGTGATTGAC GACATCACCG TTTTGATACA CGCGCAAAAA GAAGCCGCGT
1451  GGGGCGAAGT GGCAAAACGG CTGGCACACG AAATCCGCAA TCCGCTCACG
1501  CCCATCCAGC TTTCTGCCGA ACGGCTGGCG TGGAAATTGG GCGGGAAGCT
1551  GGACGAGCAN GACGCGCAAA TCCTGACACG TTCGACCGAC ACCATCATCA
1601  AACAAGTGGC GGCATTAAAA GAAATGGTCG AGGCATTCCG CAATTACNCG
1651  CGTTCCCCTT CGNCTCAATT GGAAAATCAG GATTTGAACG CCTTAATCGG
1701  CGATGTGTTG GCATTGTACG AAGCTGGTCC GTGCCGGTTT GCGGCGGAAC
1751  TTGCCGGCGA ACCGCTGATG ATGGCGGCGG ATACGACCGC CATGCGGCAG
1801  GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
1851  TGTGCCCGAA GTCAGGGTAA AATCGGAAGC GGGGCAGGAC GGACGGATTG
1901  TCCTGACAGT TTGCGACAAC GGCAAGGGGT TCGGCAGGGA AATGCTGCAC
1951  AATGCCTTCG AGCCGTATGT AACGGACAAA CCGGCTGGAA CGGGATTGNG
2001  ACTGCCCGTG GTGAAAAAAA TCATTGAAGA ACACGGCGGC CNCATCAGCC
2051  TGAGCAATCA GGATGCGGGC GGCGCGTNTG TCAGAATCAT CTTGCCAAAA
2101  ACGGTAGAAA CTTATGCGTA G
```

This encodes a protein having amino acid sequence <SEQ ID 254>:

```
  1  MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVAFS AMLLLVLSAV
 51  LARYVILLLK DRRDGVFGSQ IAKRLSGMFT LVAVLPGVFL FGVSAQFING
101  TINSWFGNDT HEALERSLNL SKSALNLAAD NALGNAIPVQ IDXIGAASLP
151  XDMGRVLEHY AGSGFAQLAL YNAASGKIEK SINPHKLDQP FPGKARWEKI
201  QQAGSVRDXE SIGGVLYAXG WLSAXTHNGR DYALFFRQPV PKGVAEDAVL
251  IEKARAXXXX LSYSKKGLQT FFLATLLIAS LLSIFLALVM ALYFARRFVE
301  PVLSLAEGAK AVAQGDFSQT RPVLRNDEFG RLTKLFNHMT EQLSIAKEAD
351  ERNRRREEAA RHYLECVLEG LTTGVVVFDE QGCLKTFNKA AEQILGMPLT
401  PLWGSSRHGW HGVSAQQSLL AEVFAAIGAA AGTDKPVHVK YAAPDDAKIL
451  LGKATVLPED NXNGVVMVID DITVLIHAQK EAAWGEVAKR LAHEIRNPLT
501  PIQLSAERLA WKLGGKLDEX DAQILTRSTD TIIKQVAALK EMVEAFRNYX
551  RSPSXQLENQ DLNALIGDVL ALYEAGPCRF AAELAGEPLM MAADTTAMRQ
601  VLHNIFKNAA EAAEEADVPE VRVKSEAGQD GRIVLTVCDN GKGFGREMLH
651  NAFEPYVTDK PAGTGLXLPV VKKIIEEHGG XISLSNQDAG GAXVRIILPK
701  TVETYA*
```

ORF64a and ORF64-1 show 96.6% identity in 706 aa overlap:

```
                    10        20        30        40        50        60
orf64a.pep  MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVAFSAMLLLVLSAVLARYVILLLK
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf64-1     MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVAFSAMLLLVLSAVLARYVILLLK
                    10        20        30        40        50        60

                    70        80        90       100       110       120
orf64a.pep  DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf64-1     DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
                    70        80        90       100       110       120

                   130       140       150       160       170       180
orf64a.pep  SKSALNLAADNALGNAIPVQIDXIGAASLPXDMGRVLEHYAGSGFAQLALYNAASGKIEK
            ||||||||||||||||:||||| ||||||| |||||||||||||||||||||||||||||
orf64-1     SKSALNLAADNALGNAVPVQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNAASGKIEK
                   130       140       150       160       170       180

                   190       200       210       220       230       240
orf64a.pep  SINPHKLDQPFPGKARWEKIQQAGSVRDXESIGGVLYAXGWLSAXTHNGRDYALFFRQPV
            ||||||||||||||||||||:||||||| |||||||||| ||||| ||||||||||||||
orf64-1     SINPHKLDQPFPGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHNGRDYALFFRQPV
                   190       200       210       220       230       240

                   250       260       270       280       290       300
```

```
orf64a.pep    PKGVAEDAVLIEKARAXXXXLSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFVE
              ||||||||||||||||    ||||||||||||||||||||||||||||||||||||||||
orf64-1       PKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFVE
                     250       260       270       280       290       300

                     310       320       330       340       350       360
orf64a.pep    PVLSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTKLFNHMTEQLSIAKEADERNRRREEAA
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf64-1       PVLSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTKLFNHMTEQLSIAKEADERNRRREEAA
                     310       320       330       340       350       360

                     370       380       390       400       410       420
orf64a.pep    RHYLECVLEGLTTGVVVFDEQGCLKTFNKAAEQILGMPLTPLWGSSRHGWHGVSAQQSLL
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf64-1       RHYLECVLEGLTTGVVVFDEQGCLKTFNKAAEQILGMPLTPLWGSSRHGWHGVSAQQSLL
                     370       380       390       400       410       420

                     430       440       450       460       470       480
orf64a.pep    AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNXNGVVMVIDDITVLIHAQK
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
orf64-1       AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNGNGVVMVIDDITVLIHAQK
                     430       440       450       460       470       480

                     490       500       510       520       530       540
orf64a.pep    EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDEXDAQILTRSTDTIIKQVAALK
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||:|||||||
orf64-1       EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDEQDAQILTRSTDTIVKQVAALK
                     490       500       510       520       530       540

                     550       560       570       580       590       600
orf64a.pep    EMVEAFRNYXRSPSXQLENQDLNALIGDVLALYEAGPCRFAAELAGEPLMMAADTTAMRQ
              ||||||||| |||| :||||||||||||||||||||||||||||||||| :|||||||||
orf64-1       EMVEAFRNYARSPSLKLENQDLNALIGDVLALYEAGPCRFAAELAGEPLTVAADTTAMRQ
                     550       560       570       580       590       600

                     610       620       630       640       650       660
orf64a.pep    VLHNIFKNAAEAAEEADVPEVRVKSEAGQDGRIVLTVCDNGKGFGREMLHNAFEPYVTDK
              ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
orf64-1       VLHNIFKNAAEAAEEADVPEVRVKSETGQDGRIVLTVCDNGKGFGREMLHNAFEPYVTDK
                     610       620       630       640       650       660

                     670       680       690       700
orf64a.pep    PAGTGLXLPVVKKIIEEHGGXISLSNQDAGGAXVRIILPKTVETYAX
              |||||| ||||||||||||| ||||||||||| |||||||||:||||
orf64-1       PAGTGLGLPVVKKIIEEHGGRISLSNQDAGGACVRIILPKTVKTYAX
                     670       680       690       700
```

<u>Homology with a predicted ORF from *N.gonorrhoeae*</u>

ORF64 shows 86.6% identity over a 387aa overlap with a predicted ORF (ORF64.ng) from *N. gonorrhoeae:*

```
orf64.pep     MRRFLPIAAICAXXLXXGLTAATGSTSSLADYFWWIVAFSAMLLLVLSAVLARYVILLLK   60
              ||||||||||||  |  ||||||||||||||||||||:||||||||||||||||||||||
orf64ng       MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVSFSAMLLLVLSAVLARYVILLLK   60

orf64.pep     DRRDGVFGSXXAKXPXXXMFTLVAXLPGVFLFGFPAQFINGTINSWFGNDTHEALERSLN  120
              |||:|||||  ||     |||||| |||:||||: |||||||||||||||||||||||||
orf64ng       DRRNGVFGSQIAKR-LSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLN  119

orf64.pep     LSKSALNLAADNALGNAVPVQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNXASGKIE  180
              ||||||:||||||::|||||||||||:||| |:|| ||||||||||||||||| ||||||
orf64ng       LSKSALDLAADNAVSNAVPVQIDLIGTASLSGNMGSVLEHYAGSGFAQLALYNAASGKIE  179

orf64.pep     KSINPHKLDQPFPGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHXGRDYALFFRQP  240
              ||||||::|||:| | :||:||::||||:||||||||||||||||||| |||||||||||
orf64ng       KSINPHQFDQPLPDKEHWEQIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYALFFRQP  239
```

```
orf64.pep    VPKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFV  300
             :|::||:|||||||||||||||||||||||||||:|||||||||||||||||||||||||
orf64ng      IPENVAQDAVLIEKARAKYAELSYSKKGLQTFFLVTLLIASLLSIFLALVMALYFARRFV  299

orf64.pep    EPVLSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTXLFNHMTEQLSIAKDADERNRRREEA  360
             ||:||||||||||||||||||||||||||||||| |||||||||||||:|||||||||||
orf64ng      EPILSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTKLFNHMTEQLSIAKEADERNRRREEA  359

orf64.pep    ARHYLECVLEGLTTGVVVFDEQGCLKTFNKAAGT         394
             |||||||||:||||||||    :| :|
orf64ng      ARHYLECVLDGLTTGVVVSYPLSCCRTAVFSTCHSSPLSYF  400
```

An ORF64ng nucleotide sequence <SEQ ID 255> was predicted to encode a protein having amino acid sequence <SEQ ID 256>:

```
  1 MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVSFS AMLLLVLSAV
 51 LARYVILLLK DRRNGVFGSQ IAKRLSGMFT LVAVLPGLFL FGISAQFING
101 TINSWFGNDT HEALERSLNL SKSALDLAAD NAVSNAVPVQ IDLIGTASLS
151 GNMGSVLEHY AGSGFAQLAL YNAASGKIEK SINPHQFDQP LPDKEHWEQI
201 QQTGSVRSLE SIGGVLYAQG WLSAGTHNGR DYALFFRQPI PENVAQDAVL
251 IEKARAKYAE LSYSKKGLQT FFLVTLLIAS LLSIFLALVM ALYFARRFVE
301 PILSLAEGAK AVAQGDFSQT RPVLRNDEFG RLTKLFNHMT EQLSIAKEAD
351 ERNRRREEAA RHYLECVLDG LTTGVVVSYP LSCCRTAVFS TCHSSPLSYF*
```

Further work revealed the complete gonococcal DNA sequence <SEQ ID 257>:

```
   1 ATGCGCCGCT TCCTACCGAT CGCAGCCATA TGCGCCGTCG TCCTGCTGTA
  51 CGGATTGACG GCGGCGACCG GCAGCACCAG TTCGCTGGCG GATTATTTCT
 101 GGTGGATAGT CTCGTTCAGC GCAATGCTGC TGCTGGTGTT GTCCGCCGTT
 151 TTGGCACGTT ATGTCATATT GCTGTTGAAA GACAGGCGCA ACGGCGTGTT
 201 CGGTTCGCAG ATTGCCAAAC GCCTTTCCGG GATGTTCACG CTGGTCGCCG
 251 TACTGCCCGG CTTGTTCCTG TTCGGCATTT CCGCGCAGTT TATCAACGGC
 301 ACGATTAATT CGTGGTTCGG CAACGACACC CACGAAGCCC TCGAACGCAG
 351 CCTTAATTTG AGCAAGTCCG CACTGGATTT GGCGGCAGAC AATGCCGTCA
 401 GCAACGCCGT TCCCGTACAG ATAGACCTCA TCGGCACCGC CTCCCTGTCG
 451 GGCAATATGG GCAGTGTGCT GGAACACTAC GCCGGCAGCG GTTTTGCCCA
 501 GCTTGCCCTG TACAATGCCG CAAGCGGGAA AATCGAAAAA AGCATCAATC
 551 CGCACCAATT CGACCAGCCG CTTCCCGACA AAGAACATTG GGAACAGATT
 601 CAGCAGACCG GTTCGGTTCG GAGTTTGGAA AGCATAGGCG GCGTATTGTA
 651 CGCGCAGGGA TGGTTGTCGG CAGGTACGCA CAACGGGCGC GATTACGCGC
 701 TGTTCTTCCG CCAGCCGATT CCCGAAAATG TGGCACAGGA TGCCGTTCTG
 751 ATTGAAAAGG CGCGGGCGAA ATATGCCGAA TTGAGTTACA GCAAAAAAGG
 801 TTTGCAGACC TTTTTTCTGG TAACCCTGCT GATTGCCTCG CTGCTGTCGA
 851 TTTTTCTTGC GCTGGTAATG GCACTGTATT TTGCCCGCCG TTTCGTCGAA
 901 CCCATTCTGT CGCTTGCCGA GGGCGCAAAG GCGGTGGCGC AGGGTGATTT
 951 CAGCCAGACG CGCCCCGTAT TGCGCAACGA CGAGTTCGGA CGTTTGACCA
1001 AGCTGTTCAA CCATATGACC GAGCAGCTTT CCATCGCCAA AGAAGCAGAC
1051 GAACGCAACC GCCGGCGCGA GGAAGCCGCC CGTCACTACC TCGAGTGCGT
1101 GTTGGATGGG TTGACTACCG GTGTGGTGGT GTTTGACGAA AAAGGCCGTT
1151 TGAAAACCTT CAACAAGGCG GCGGAACAGA TTTTGGGGAT GCCGCTCGCC
1201 CCCCTGTGGG GCAGCAGCCG GCACGGTTGG CACGGCGTTT CGGCGCAGCA
1251 GTCCCTGCTT GCCGAAGTGT TtgccgccAT CGGTGCGGCG GCAGGTACGG
1301 ACAAACCGGT CCAGGTGGAA TATGCCGCGC CGGACGATGC CAAAATCCTG
1351 CTGGGCAAGG CGACGGTATT GCCCGAAGAC AACGGCAACG GCGTGGTGAT
1401 GGTGATTGAC GACATCACCG TGCTGATACG CGCGCAAAAA GAAGCCGCGT
1451 GGGGTGAAGT GGCGAAGCGG CTGGCACACG AAATCCGCAA TCCGCTCACG
1501 CCCATCCAGC TTTCCGCCGA ACGGCTGGCG TGGAAATTGG GCGGGAAGCT
1551 GGACGATCAG GACGCGCAAA TCCTGACGCG TtcgACCGAC ACCATCATCA
1601 AACAGgtggc gGCGTTAAAA GAAATGGTCG AGGCATTCCG CAATTACGCG
1651 CGCGCCCCTT CGCTCAAACT GGAAAATCAG GATTTGAACG CCTTAATCGG
1701 CGATGTTTTG GCCCTGTACG AAGCCGGCCC GTGCCGGTTT GAGGCGGAAC
1751 TTGCCGGCGA ACCGCTGATG ATGGCGGCGG ATACGACCGC CATGCGGCAG
1801 GTGCTGCACA ATATTTTCAA AAATGCCGCC GAAGCGGCGG AAGAAGCCGA
1851 TATGCCCGAA GTCAGGGTAA AATCGGAAAC GGGGCAGGAC GGACGGATTG
1901 TCCTGACGGT TTGCGACAAC GGCAAGGGAT TCGGCAAGGA AATGCTGCAC
1951 AATGCTTTCG AGCCGTATGT GACGGATAAG CCGGCGGGAA CGGGACTGGG
2001 TCTGCCTGTA GTGAAAAAAA TCATTGGAGA ACACGGCGGC CGCATCAGCC
2051 TGAGCAATCA GGATGCGGGT GGGGCGTGTG TCAGAATCAT CTTGCCAAAA
2101 ACGGTAGAAA CTTATGCGTA G
```

This corresponds to the amino acid sequence <SEQ ID 258; ORF64ng-1>:

```
  1  MRRFLPIAAI CAVVLLYGLT AATGSTSSLA DYFWWIVSFS AMLLLVLSAV
 51  LARYVILLLK DRRNGVFGSQ IAKRLSGMFT LVAVLPGLFL FGISAQFING
101  TINSWFGNDT HEALERSLNL SKSALDLAAD NAVSNAVPVQ IDLIGTASLS
151  GNMGSVLEHY AGSGFAQLAL YNAASGKIEK SINPHQFDQP LPDKEHWEQI
201  QQTGSVRSLE SIGGVLYAQG WLSAGTHNGR DYALFFRQPI PENVAQDAVL
251  IEKARAKYAE LSYSKKGLQT FFLVTLLIAS LLSIFLALVM ALYFARRFVE
301  PILSLAEGAK AVAQGDFSQT RPVLRNDEFG RLTKLFNHMT EQLSIAKEAD
351  ERNRRREEAA RHYLECVLDG LTTGVVVFDE KGRLKTFNKA AEQILGMPLA
401  PLWGSSRHGW HGVSAQQSLL AEVFAAIGAA AGTDKPVQVE YAAPDDAKIL
451  LGKATVLPED NGNGVVMVID DITVLIRAQK EAAWGEVAKR LAHEIRNPLT
501  PIQLSAERLA WKLGGKLDDQ DAQILTRSTD TIIKQVAALK EMVEAFRNYA
551  RAPSLKLENQ DLNALIGDVL ALYEAGPCRF EAELAGEPLM MAADTTAMRQ
601  VLHNIFKNAA EAAEEADMPE VRVKSETGQD GRIVLTVCDN GKGFGKEMLH
651  NAFEPYVTDK PAGTGLGLPV VKKIIGEHGG RISLSNQDAG GACVRIILPK
701  TVETYA*
```

ORF64ng-1 and ORF64-1 show 93.8% identity in 706 aa overlap:

```
                        10        20        30        40        50        60
orf64ng-1.pep   MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVSFSAMLLLVLSAVLARYVILLLK
                ||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
orf64-1         MRRFLPIAAICAVVLLYGLTAATGSTSSLADYFWWIVAFSAMLLLVLSAVLARYVILLLK
                        10        20        30        40        50        60

                        70        80        90       100       110       120
orf64ng-1.pep   DRRNGVFGSQIAKRLSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLNL
                |||:|||||||||||||||||||||||:||||:|||||||||||||||||||||||||||
orf64-1         DRRDGVFGSQIAKRLSGMFTLVAVLPGVFLFGVSAQFINGTINSWFGNDTHEALERSLNL
                        70        80        90       100       110       120

                       130       140       150       160       170       180
orf64ng-1.pep   SKSALDLAADNAVSNAVPVQIDLIGTASLSGNMGSVLEHYAGSGFAQLALYNAASGKIEK
                |||||:||||||::|||||||||||:||| |:|| |||||||||||||||||||||||||
orf64-1         SKSALNLAADNALGNAVPVQIDLIGAASLPGDMGRVLEHYAGSGFAQLALYNAASGKIEK
                       130       140       150       160       170       180

                       190       200       210       220       230       240
orf64ng-1.pep   SINPHQFDQPLPDKEHWEQIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYALFFRQPI
                |||||::|||:| | :||:||::||||:|||||||||||||||||||||||||||||||:
orf64-1         SINPHKLDQPFPGKARWEKIQRAGSVRDLESIGGVLYAQGWLSAGTHNGRDYALFFRQPV
                       190       200       210       220       230       240

                       250       260       270       280       290       300
orf64ng-1.pep   PENVAQDAVLIEKARAKYAELSYSKKGLQTFFLVTLLIASLLSIFLALVMALYFARRFVE
                |::||:|||||||||||||||||||||||||||:||||||||||||||||||||||||||
orf64-1         PKGVAEDAVLIEKARAKYAELSYSKKGLQTFFLATLLIASLLSIFLALVMALYFARRFVE
                       250       260       270       280       290       300

                       310       320       330       340       350       360
orf64ng-1.pep   PILSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTKLFNHMTEQLSIAKEADERNRRREEAA
                |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf64-1         PVLSLAEGAKAVAQGDFSQTRPVLRNDEFGRLTKLFNHMTEQLSIAKEADERNRRREEAA
                       310       320       330       340       350       360

                       370       380       390       400       410       420
orf64ng-1.pep   RHYLECVLDGLTTGVVVFDEKGRLKTFNKAAEQILGMPLAPLWGSSRHGWHGVSAQQSLL
                ||||||||:|||||||||||:| ||||||||||||||||:||||||||||||||||||||
orf64-1         RHYLECVLEGLTTGVVVFDEQGCLKTFNKAAEQILGMPLTPLWGSSRHGWHGVSAQQSLL
                       370       380       390       400       410       420

                       430       440       450       460       470       480
orf64ng-1.pep   AEVFAAIGAAAGTDKPVQVEYAAPDDAKILLGKATVLPEDNGNGVVMVIDDITVLIRAQK
                |||||||||||||||||:|:||||||||||||||||||||||||||||||||||||:|||
orf64-1         AEVFAAIGAAAGTDKPVHVKYAAPDDAKILLGKATVLPEDNGNGVVMVIDDITVLIHAQK
                       430       440       450       460       470       480

                       490       500       510       520       530       540
orf64ng-1.pep   EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDDQDAQILTRSTDTIIKQVAALK
                ||||||||||||||||||||||||||||||||||||||:|||||||||||||:|||||||
```

```
orf64-1         EAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDEQDAQILTRSTDTIVKQVAALK
                        490       500       510       520       530       540

                        550       560       570       580       590       600
orf64ng-1.pep   EMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFEAELAGEPLMMAADTTAMRQ
                ||||||||||:||||||||||||||||||||||||||||| |||||||| :|||||||||
orf64-1         EMVEAFRNYARSPSLKLENQDLNALIGDVLALYEAGPCRFAAELAGEPLTVAADTTAMRQ
                        550       560       570       580       590       600

                        610       620       630       640       650       660
orf64ng-1.pep   VLHNIFKNAAEAAEEADMPEVRVKSETGQDGRIVLTVCDNGKGFGKEMLHNAFEPYVTDK
                |||||||||||||||||:||||||||||||||||||||||||||||:|||||||||||||
orf64-1         VLHNIFKNAAEAAEEADVPEVRVKSETGQDGRIVLTVCDNGKGFGREMLHNAFEPYVTDK
                        610       620       630       640       650       660

                        670       680       690       700
orf64ng-1.pep   PAGTGLGLPVVKKIIGEHGGRISLSNQDAGGACVRIILPKTVETYAX
                |||||||||||||| |||||||||||||||||||||||||||:||||
orf64-1         PAGTGLGLPVVKKIIEEHGGRISLSNQDAGGACVRIILPKTVKTYAX
                        670       680       690       700
```

Furthermore, ORF64ng-1 shows significant homology to a protein from *A.caulinodans*:

```
sp|Q04850|NTRY_AZOCA NITROGEN REGULATION PROTEIN NTRY >gi|77479|pir||S18624 ntrY
protein - Azorhizobium caulinodans >gi|38737 (X63841) NtrY gene product
[Azorhizobium caulinodans] Length = 771
 Score =  218 bits (550), Expect = 7e-56
 Identities = 195/720 (27%), Positives = 320/720 (44%), Gaps = 58/720 (8%)

Query: 7   IAAICAVVLLYGLTAATGSTSSLADYFWWIXXXXXXXXXXXXXXXXXRYVILLLKDRRNGV 66
            I+A+   ++L GLT     +     +                     R +  + K R  G
Sbjct: 35  ISALATFLILMGLTPVVPTHQVVIS----VLLVNAAAVLILSAMVGREIWRIAKARARGR 90

Query: 67  FGSQIAKRLSGMFTLVAVLPGLFLFGISAQFINGTINSWFGNDTHEALERSLNLSKSALD 126
              +++  R+ G+F +V+V+P + +  +++  ++  ++ WF   T E +  S++++++ +
Sbjct: 91  AAARLHIRIVGLFAVVSVVPAILVAVVASLTLDRGLDRWFSMRTQEIVASSVSVAQTYVR 150

Query: 127 LAADNAVSNAVPVQIDLIGTASLSGNMGSVLEHYAG--SGFAQLALYNAASGKIEKSINP 184
              A N  + + + DL   S+             Y G  S F Q+     AA   +  ++
Sbjct: 151 EHALNIRGDILAMSADLTRLKSV----------YEGDRSRFNQILTAQAALRNLPGAMLI 200

Query: 185 HQFDQPLPDKEHWEQIQQTGSVRSLESIGGVLYAQGWLSAGTHNGRDYA----------- 233
             + D  + ++ +   I +   V +  +IG      Q  +     N  DY
Sbjct: 201 RR-DLSVVERAN-VNIGREFIVPANLAIGDATPDQPVIYLP--NDADYVAAVVPLKDYDD 256

Query: 234 --LFFRQPIPENVAQDAVLIEKARAKYAELSYSKKGLQTFFLVTXXXXXXXXXXXXXXVMA 291
      L+  + I   V       ++  A Y   L    + G+Q   F +              +
Sbjct: 257 LYLYVARLIDPRVIGYLKTTQETLADYRSLEERRFGVQVAFALMYAVITLIVLLSAVWLG 316

Query: 292 LYFARRFVEPILSLAEGAKAVAQGDFSQTRPVLRND-EFGRLTKLFNHMTEQLSIXXXXX 350
            L F++  V PI  L   A  VA+G+      P+ R + +   L + FN MT +L
Sbjct: 317 LNFSKWLVAPIRRLMSAADHVAEGNLDVRVPIYRAEGDLASLAETFNKMTHELRSQREAI 376

Query: 351 XXXXXXXXXXXHYLECVLDGLTTGVVVFDEKGRLKTFNKAAEQILGMPLAPLWGSSRHGW 410
                        + E VL G+  GV+  D + R+   N++AE++LG  L+ +     RH
Sbjct: 377 LTARDQIDSRRRFTEAVLSGVGAGVIGLDSQERITILNRSAERLLG--LSEVEALHRHLA 434

Query: 411 HGVSAQQSLLAEVFXXXXXXXXXTDKPVQVEYAAPDDAKILLGKATVLPEDNG---NGVVM 467
              V     LL E                + VQ         D +  +      V  E +   +G V+
Sbjct: 435 EVVPETAGLLEEA------EHARQRSVQGNITLTRDGRERVFAVRVTTEQSPEAEHGWVV 488

Query: 468 VIDDITVLIRAQKEAAWGEVAKRLAHEIRNPLTPIQLSAERLAWKLGGKLDDQDAQILTR 527
             +DDIT  LI AQ+ +AW +VA+R+AHEI+NPLTPIQLSAERL   K G   +  QD +I  +
Sbjct: 489 TLDDITELISAQRTSAWADVARRIAHEIKNPLTPIQLSAERLKRKFGRHV-TQDREIFDQ 547

Query: 528 STDTIIKQVAALKEMVEAFRNYARAPSLKLENQDLNALIGDVLALYEAGPCRFEAELAGE 587
             TDTII+QV   +  MV+  F ++AR  P    +++QD++ +I    + L   G        +
Sbjct: 548 CTDTIIRQVGDIGRMVDEFSSFARMPKPVVDSQDMSEIIRQTVFLMRVGHPEVVFDSEVP 607

Query: 588 PLMMAA-DTTAMRQVLHNIFKNXXXXXXXXXDMPEVRVK-------SETGQDGRIVLTVCD 639
            P M A  D   + Q L NI KN               P+VR +         +  G+D  +V+ + D
Sbjct: 608 PAMPARFDRRLVSQALTNILKNAAEAIEAVP-PDVRGQGRIRVSANRVGED--LVIDIID 664
```

```
Query: 640 NGKGFGKEMLHNAFEPYVTDKPAGTGLGLPVVKKIIGEHGGRISLSNQDAG-GACVRIIL 698
           NG G  +E  +    EPYVT +  GTGLGL +V KI+ EHGG I L++   G GA +R+ L
Sbjct: 665 NGTGLPQESRNRLLEPYVTTREKGTGLGLAIVGKIMEEHGGGIELNDAPEGRGAWIRLTL 724
```

Based on this analysis, including the presence of a putative leader sequence (double-underlined) and several putative transmembrane domains (single-underlined) in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 31

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 259>:

```
  1  ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
 51  GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101  CTTTCCAAAT TTTCGGCATC CACACCACTT GGGGCGCATT TTCCTTTCCC
151  TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT
201  GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCCGCCCTT TTGCTTTCCT
251  ACGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACAGG CTTGGGCGCG
301  CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCCTTAG CCAGCTTTGC
351  CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTTGTATTC AACAAATTAC
401  GCCGTCTGAA AGCGTGGTGG ATTGCACCGA ACGCATCAAC CGTCATCGGG
451  CACGCGTTGG ATACG...
```

This corresponds to the amino acid sequence <SEQ ID 260; ORF66>:

```
  1  MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFQIFGI HTTWGAFSFP
 51  FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
101  LSEFNTFVGR IALASFAAYA IGQILDIFVF NKLRRLKAWW IAPNASTVIG
151  HALDT...
```

Further work revealed the complete nucleotide sequence <SEQ ID 261>:

```
  1  ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
 51  GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101  CTTTCCAAAT TTTCGGCATC CACACCACTT GGGGCGCATT TTCCTTTCCC
151  TTCATCTTCC TTGCCACCGA CCTGACCGTC CGCATTTTCG GTTCTCACTT
201  GGCACGGCGG ATTATCTTTT GGGTGATGTT CCCCGCCCTT TTGCTTTCCT
251  ACGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACAGG CTTGGGCGCG
301  CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCCTTAG CCAGCTTTGC
351  CGCCTACGCG ATCGGACAAA TCCTTGATAT TTTTGTATTC AACAAATTAC
401  GCCGTCTGAA AGCGTGGTGG ATTGCACCGA CCGCATCAAC CGTCATCGGC
451  AACGCCTTGG ATACGCTGGT ATTTTTCGCC GTTGCCTTCT ACGCAAGCAG
501  CGATGGATTT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
551  TGTTCAAACT TACCGTCTGC ACCCTCTTCT TCCTGCCCGC CTACGGCGTG
601  ATACTGAATC TGCTGACGAA AAAACTGACA ACCCTGCAAA CCAAACAGGC
651  GCAAGACCGC CCCGCGCCCT CGCTGCAAAA TCCGTAA
```

This corresponds to the amino acid sequence <SEQ ID 262; ORF66-1>:

```
  1  MYAFTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFQIFGI HTTWGAFSFP
 51  FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
101  LSEFNTFVGR IALASFAAYA IGQILDIFVF NKLRRLKAWW IAPTASTVIG
151  NALDTLVFFA VAFYASSDGF MAANWQGIAF VDYLFKLTVC TLFFLPAYGV
201  ILNLLTKKLT TLQTKQAQDR PAPSLQNP*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with the hypothetical protein o221 of *E. coli* (accession number P37619)

ORF66 and o221 protein show 67% aa identity in 155aa overlap:

```
orf66  1   MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV 60
           M  F+  Q+ KALF L LFH+L+I +SNYLVQ P  I G HTTWGAFSFPFIFLATDLTV
o221   1   MNVFSQTQRYKALFWLSLFHLLVITSSNYLVQLPVSILGFHTTWGAFSFPFIFLATDLTV 60

orf66  61  RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA 120
           RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA
o221   61  RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQGFGALAHFNLFVARIATASFMAYA 120

orf66  121 IGQILDIFVFNKLRRLKAWWIAPNASTVIGHALDT 155
           +GQILD+ VFN+LR+ + WW+AP AST+ G+  DT
o221   121 LGQILDVHVFNRLRQSRRWWLAPTASTLFGNVSDT 155
```

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF66 shows 96.1% identity over a 155aa overlap with an ORF (ORF66a) from strain A of *N. meningitidis:*

```
                       10        20        30        40        50        60
orf66.pep      MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV
               |||||||||||||| ||||||||||||||||||||||| ||||||||||||||||||||||
orf66a         MYAFTAAQQQKALFWLVLFHILIIAASNYLVQFPFQISGIHTTWGAFSFPFIFLATDLTV
                       10        20        30        40        50        60

                       70        80        90       100       110       120
orf66.pep      RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf66a         RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA
                       70        80        90       100       110       120

                      130       140       150
orf66.pep      IGQILDIFVFNKLRRLKAWWIAPNASTVIGHALDT
               :||||||||||||||||||||:||:||||||:||||
orf66a         LGQILDIFVFNKLRRLKAWWVAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGIAF
                      130       140       150       160       170       180

orf66a         VDYLFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX
                      190       200       210       220
```

The complete length ORF66a nucleotide sequence <SEQ ID 263> is:

```
  1 ATGTACGCAT TTACCGCCGC ACAGCAACAG AAGGCACTCT TCTGGCTGGT
 51 GCTTTTTCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101 CCTTCCAAAT TTCCGGCATC CACACCACTT GGGGCGCGTT TTCCTTTCCC
151 TTCATCTTCC TCGCCACCGA CCTGACCGTC CGCATTTTCG GTTCGCACTT
201 GGCACGGCGG ATTATCTTTT GGGTCATGTT CCCCGCCCTT TTGCTTTCCT
251 ACGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACGGG CTTGGGCGCG
301 CTGTCCGAAT TCAACACCTT TGTCGGACGC ATCGCGCTGG CAAGTTTTGC
351 CGCCTACGCG CTCGGACAAA TCCTTGATAT TTTTGTGTTC AACAAATTAC
401 GCCGTCTGAA AGCGTGGTGG GTTGCCCCGA CTGCATCAAC CGTCATCGGC
451 AACGCCTTAG ATACGTTGGT ATTTTTCGCC GTTGCCTTCT ACGCAAGCAG
501 CGATGGATTT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
551 TGTTCAAACT CACCGTCTGC GGTCTGTTTT TCCTGCCCGC CTACGGCGTG
601 ATTCTGAATC TGCTGACGAA AAAACTGACG ACCCTGCAAA CCAAACAGGC
651 GCAAGACCGC CCCGCGCCCT CGCTGCAAAA TCCGTAA
```

This encodes a protein having amino acid sequence <SEQ ID 264>:

```
  1 MYAFTAAQQQ KALFWLVLFH ILIIAASNYL VQFPFQISGI HTTWGAFSFP
 51 FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
101 LSEFNTFVGR IALASFAAYA LGQILDIFVF NKLRRLKAWW VAPTASTVIG
151 NALDTLVFFA VAFYASSDGF MAANWQGIAF VDYLFKLTVC GLFFLPAYGV
201 ILNLLTKKLT TLQTKQAQDR PAPSLQNP*
```

ORF66a and ORF66-1 show 97.8% identity in 228 aa overlap:

```
                       10        20        30        40        50        60
orf66a.pep     MYAFTAAQQQKALFWLVLFHILIIAASNYLVQFPFQISGIHTTWGAFSFPFIFLATDLTV
               |||||||||||||| ||||||||||||||||||||||| ||||||||||||||||||||||
orf66-1        MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV
```

```
                    10        20        30        40        50        60

                    70        80        90       100       110       120
orf66a.pep   RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf66-1      RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA
                    70        80        90       100       110       120

                   130       140       150       160       170       180
orf66a.pep   LGQILDIFVFNKLRRLKAWWVAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGIAF
             :|||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
orf66-1      IGQILDIFVFNKLRRLKAWWIAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGIAF
                   130       140       150       160       170       180

                   190       200       210       220      229
orf66a.pep   VDYLFKLTVCGLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX
             |||||||||| ||||||||||||||||||||||||||||||||||||||
orf66-1      VDYLFKLTVCTLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX
                   190       200       210       220
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF66shows 94.2% identity over a 155aa overlap with a predicted ORF (ORF66.ng) from *N. gonorrhoeae:*

```
orf66.pep    MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV   60
             |||:|||||||||||||||||||||||||||||||:||||||||||||||||||||||||
orf66ng      MYALTAAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV   60

orf66.pep    RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA  120
             |||||||||||||||||||| |||||||||||||||||||| |:||||||||||||||||
orf66ng      RIFGSHLARRIIFWVMFPALSLSYVFSVLFHNGSWTGLGAPSQFNTFVGRIALASFAAYA  120

orf66.pep    IGQILDIFVFNKLRRLKAWWIAPNASTVIGHALDT                           155
             :||||||||:||||||||||||| ||||||:||||
orf66ng      LGQILDIFVFDKLRRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGIAF  180
```

The complete length ORF66ng nucleotide sequence <SEQ ID 265> is:

```
  1  ATGTACGCAT TGACCGCCGC ACAGCAACAG AAGGCACTCT TCCGGCTGGT
 51  GCTTTTCCAT ATCCTCATCA TCGCCGCCAG CAACTATCTG GTGCAGTTCC
101  CCTTCCGGAT TTTCGGCATC CACACCACTT GGGGCGCGTT TTCCTTTCCC
151  TTCATCTTCC TCGCCACCGA CCTGACCGTC CGCATTTTCG GTTCGCACTT
201  GGCGCGGCGG ATTATCTTTT GGGTGATGTT CCCCGCCCTT ttgCTTTcat
251  aCGTCTTTTC CGTTTTGTTC CACAACGGCA GTTGGACGGG CTTGGGCGCG
301  ctgTCCCAAT TCAACACCTT TGTCGGACGC ATCGCGCTGG CAAGTTTTGC
351  CGCCTACGCG CTCGGACAAA TCCTTGATAT TTTCGTATTC GACAAATTAC
401  GCCGTCTGAA AGCGTGGTGG ATTGCCCCGG CCGCATCAAC CGTCATCGGC
451  AATGCACTGG ACACGTTAGT ATTTTTTGCC GTTGCCTTTT ACGCAAGCAG
501  CGATGAATTT ATGGCGGCAA ACTGGCAGGG CATCGCTTTT GTCGATTACC
551  TGTTCAAACT TACCGTCTGC ACCCTCTTCT TCCTGCCCGC CTACGGCGTG
601  ATACTGAATC TGCTGACGAA AAAACTGACG GCCCTGCAAA CCAAACAGGC
651  GCAAGACCGC CCCGTGCCCT CGCTGCAAAA TCCGTAA
```

This encodes a protein having amino acid sequence <SEQ ID 266>:

```
  1  MYALTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFRIFGI HTTWGAFSFP
 51  FIFLATDLTV RIFGSHLARR IIFWVMFPAL SLSYVFSVLF HNGSWTGLGA
101  PSQFNTFVGR IALASFAAYA LGQILDIFVF DKLRRLKAWW IAPAASTVIG
151  NALDTLVFFA VAFYASSDEF MAANWQGIAF VDYLFKLTVC TLFFLPAYGV
201  ILNLLTKKLT ALQTKQAQDR PVPSLQNP*
```

An alternative annotated sequence is:

```
  1  MYALTAAQQQ KALFRLVLFH ILIIAASNYL VQFPFRIFGI HTTWGAFSFP
 51  FIFLATDLTV RIFGSHLARR IIFWVMFPAL LLSYVFSVLF HNGSWTGLGA
101  LSQFNTFVGR IALASFAAYA LGQILDIFVF DKLRRLKAWW IAPAASTVIG
151  NALDTLVFFA VAFYASSDEF MAANWQGIAF VDYLFKLTVC TLFFLPAYGV
201  ILNLLTKKLT ALQTKQAQDR PVPSLQNP*
```

ORF66ng and ORF66-1 show 96.1% identity in 228 aa overlap:

```
orf66-1.pep  MYAFTAAQQQKALFRLVLFHILIIAASNYLVQFPFQIFGIHTTWGAFSFPFIFLATDLTV  60
             |||:||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
orf66ng      MYALTAAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV  60

orf66-1.pep  RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSEFNTFVGRIALASFAAYA 120
             |||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
orf66ng      RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120

orf66-1.pep  IGQILDIFVFNKLRRLKAWWIAPTASTVIGNALDTLVFFAVAFYASSDGFMAANWQGIAF 180
             :|||||||||:|||||||||||:|||||||||||||||||||||||||| |||||||||||
orf66ng      LGQILDIFVFDKLRRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGIAF 180

orf66-1.pep  VDYLFKLTVCTLFFLPAYGVILNLLTKKLTTLQTKQAQDRPAPSLQNPX   229
             |||||||||||||||||||||||||||||||:|||||||||||:|||||||
orf66ng      VDYLFKLTVCTLFFLPAYGVILNLLTKKLTALQTKQAQDRPVPSLQNPX   229
```

Furthermore, ORF66ng shows significant homology with an *E.coli* ORF:

```
 sp|P37619|YHHQ_ECOLI HYPOTHETICAL 25.3 KD PROTEIN IN FTSY-NIKA INTERGENIC
REGION (O221)
>gi|1073495|pir||S47690 hypothetical protein o221 - Escherichia coli >gi|466607
(U00039) No definition line found [Escherichia coli] >gi|1789882 (AE000423)
hypothetical 25.3 kD protein in ftsY-nikA intergenic region [Escherichia coli]
Length = 221
 Score =  273 bits (692), Expect = 5e-73
 Identities = 132/203 (65%), Positives = 155/203 (76%)

Query: 1    MYALTAAQQQKALFRLVLFHILIIAASNYLVQFPFRIFGIHTTWGAFSFPFIFLATDLTV 60
            M   +  Q+ KALF L LFH+L+I +SNYLVQ P   I G HTTWGAFSFPFIFLATDLTV
Sbjct: 1    MNVFSQTQRYKALFWLSLFHLLVITSSNYLVQLPVSILGFHTTWGAFSFPFIFLATDLTV 60

Query: 61   RIFGSHLARRIIFWVMFPALLLSYVFSVLFHNGSWTGLGALSQFNTFVGRIALASFAAYA 120
            RIFG+ LARRIIF VM PALL+SYV S LF+ GSW G GAL+ FN FV RIA ASF AYA
Sbjct: 61   RIFGAPLARRIIFAVMIPALLISYVISSLFYMGSWQGFGALAHFNLFVARIATASFMAYA 120

Query: 121  LGQILDIFVFDKLRRLKAWWIAPAASTVIGNALDTLVFFAVAFYASSDEFMAANWQGIAF 180
            LGQILD+ VF++LR+ + WW+AP AST+ GN  DTL FF +AF+ S D FMA +W  IA
Sbjct: 121  LGQILDVHVFNRLRQSRRWWLAPTASTLFGNVSDTLAFFFIAFWRSPDAFMAEHWMEIAL 180

Query: 181  VDYLFKLTVCTLFFLPAYGVILN 203
            VDY FK+ +   +FFLP YGV+LN
Sbjct: 181  VDYCFKVLISIVFFLPMYGVLLN 203
```

Based on this analysis, including the homology with the *E.coli* protein and the presence of several putative transmembrane domains in the gonococcal protein, it is predicted that these proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 32

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 267>:

```
  1  ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
 51  AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAAyGCA GTmwrAATAT
101  CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
151  GTACCTAAAA ATAGTAAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
201  TTTAACACAC AyyCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
251  CCGCCAGCGT ATCCCGCGCC GGCGTATTGG CGGGGGTCGG CAAACTTGCC
301  CGCTTAGgCG CGAAATTCAG CACAAGGGCG GTtCCCTATG TCGGAACAGC
351  CcTTTTAGCC CACGACGTAT ACGAAAcTTT CAAAGAAGAC ATACAGGCAC
401  GAGGCTACCA ATACGACCCC GAAACCGACA AATTTGTAAA AGGCTACGAA
451  TATAGTAATT GCCTTTGGTA CGAAGACAAA AGACGTATTA ATAGAACCTA
```

```
501  TGGCTGCTAC GGCGTTGAT..
```

This corresponds to the amino acid sequence <SEQ ID 268; ORF72>:

```
  1  MVIKYTNLNF AKLSIIAILM MYSFEANANA VXISETVSVD TGQGAKIHKF
 51  VPKNSKTYSS DLIKTVDLTH XPTGAKARIN AKITASVSRA GVLAGVGKLA
101  RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFVKGYE
151  YSNCLWYEDK RRINRTYGCY GVD..
```

Further work revealed the complete nucleotide sequence <SEQ ID 269>:

```
  1  ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
 51  AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
101  CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
151  GTACCTAAAA ATAGTAAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
201  TTTAACACAC ATCCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
251  CCGCCAGCGT ATCCCGCGCC GGCGTATTGG CGGGGGTCGG CAAACTTGCC
301  CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTTCCCTATG TCGGAACAGC
351  CCTTTTAGCC CACGACGTAT ACGAAACTTT CAAAGAAGAC ATACAGGCAC
401  GAGGCTACCA ATACGACCCC GAAACCGACA AATTTGCAAA GGTCTCAGGC
451  TAA
```

This corresponds to the amino acid sequence <SEQ ID 270; ORF72-1>:

```
  1  MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF
 51  VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAGVGKLA
101  RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFAKVSG
151  *
```

Computer analysis of this amino acid sequence gave the following results:

<u>Homology with a predicted ORF from *N.meningitidis* (strain A)</u>

ORF72 shows 98.0% identity over a 147aa overlap with an ORF (ORF72a) from strain A of *N. meningitidis:*

```
                    10        20        30        40        50        60
orf72.pep   MVIKYTNLNFAKLSIIAILMMYSFEANANAVXISETVSVDTGQGAKIHKFVPKNSKTYSS
            |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
orf72a      MVIKYTNLNFAKLSIIAILMMYSFEANANAVKISETVSVDTGQGAKIHKFVPKNSKTYSS
                    10        20        30        40        50        60

                    70        80        90       100       110       120
orf72.pep   DLIKTVDLTHXPTGAKARINAKITASVSRAGVLAGVGKLARLGAKFSTRAVPYVGTALLA
            |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
orf72a      DLIKTVDLTHIPTGAKARINAKITASVSRAGVLAGVGKLARLGAKFSTRAVPYVGTALLA
                    70        80        90       100       110       120

                   130       140       150       160       170
orf72.pep   HDVYETFKEDIQARGYQYDPETDKFVKGYEYSNCLWYEDKRRINRTYGCYGVD
            ||||||||||||||||||||||||:|
orf72a      HDVYETFKEDIQARGYQYDPETDKFAKVSGX
                   130       140       150
```

The complete length ORF72a nucleotide sequence <SEQ ID 271> is:

```
  1  ATGGTCATAA AATATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
 51  AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
101  CTGAAACTGT TTCAGTTGAT ACCGGACAAG GTGCGAAAAT TCATAAGTTT
151  GTACCTAAAA ATAGTAAAAC TTATTCATCT GATTTAATAA AAACGGTAGA
201  TTTAACACAC ATCCCTACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
251  CCGCCAGCGT ATCCCGCGCC GGCGTATTGG CGGGGGTCGG CAAACTTGCC
301  CGCTTAGGCG CGAAATTCAG CACAAGGGCG GTTCCCTATG TCGGAACAGC
351  CCTTTTAGCC CACGACGTAT ACGAAACTTT CAAAGAAGAC ATACAGGCAC
401  GAGGCTACCA ATACGACCCC GAAACCGACA AATTTGCAAA GGTCTCAGGC
451  TAA
```

This encodes a protein having amino acid sequence <SEQ ID 272>:

```
  1  MVIKYTNLNF AKLSIIAILM MYSFEANANA VKISETVSVD TGQGAKIHKF
 51  VPKNSKTYSS DLIKTVDLTH IPTGAKARIN AKITASVSRA GVLAGVGKLA
101  RLGAKFSTRA VPYVGTALLA HDVYETFKED IQARGYQYDP ETDKFAKVSG
151  *
```

ORF72a and ORF72-1 show 100.0% identity in 150 aa overlap:

```
                     10        20        30        40        50        60
orf72a.pep   MVIKYTNLNFAKLSIIAILMMYSFEANANAVKISETVSVDTGQGAKIHKFVPKNSKTYSS
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf72-1      MVIKYTNLNFAKLSIIAILMMYSFEANANAVKISETVSVDTGQGAKIHKFVPKNSKTYSS
                     10        20        30        40        50        60

                     70        80        90       100       110       120
orf72a.pep   DLIKTVDLTHIPTGAKARINAKITASVSRAGVLAGVGKLARLGAKFSTRAVPYVGTALLA
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf72-1      DLIKTVDLTHIPTGAKARINAKITASVSRAGVLAGVGKLARLGAKFSTRAVPYVGTALLA
                     70        80        90       100       110       120

                    130       140       150
orf72a.pep   HDVYETFKEDIQARGYQYDPETDKFAKVSGX
             |||||||||||||||||||||||||||||||
orf72-1      HDVYETFKEDIQARGYQYDPETDKFAKVSGX
                    130       140       150
```

<u>Homology with a predicted ORF from *N.gonorrhoeae*</u>

ORF72 shows 89% identity over a 173aa overlap with a predicted ORF (ORF72.ng) from *N. gonorrhoeae:*

```
orf72.pep    MVIKYTNLNFAKLSIIAILMMYSFEANANAVXISETVSVDTGQGAKIHKFVPKNSKTYSS   60
             || |:||||||||||||||||||||||||||| ||||:|||||||||:||||||:|: |||
orf72ng      MVTKHTNLNFAKLSIIAILMMYSFEANANAVKISETLSVDTGQGAKVHKFVPKSSNIYSS   60

orf72.pep    DLIKTVDLTHXPTGAKARINAKITASVSRAGVLAGVGKLARLGAKFSTRAVPYVGTALLA  120
             || |:||||| ||||||||||||||||||||||:|||||:| ||||:|||||||||||||
orf72ng      DLTKAVDLTHIPTGAKARINAKITASVSRAGVLSGVGKLVRQGAKFGTRAVPYVGTALLA  120

orf72.pep    HDVYETFKEDIQARGYQYDPETDKFVKGYEYSNCLWYEDKRRINRTYGCYGVD          173
             ||||||||||||||| :||||||||||||||:|||||||:|||||||||||||
orf72ng      HDVYETFKEDIQARGCRYDPETDKFVKGYEYANCLWYEDERRINRTYGCYGVDSSIMRLM  180
```

An ORF72ng nucleotide sequence <SEQ ID 273> was predicted to encode a protein having amino acid sequence <SEQ ID 274>:

```
  1  MVTKHTNLNF AKLSIIAILM MYSFEANANA VKISETLSVD TGQGAKVHKF
 51  VPKSSNIYSS DLTKAVDLTH IPTGAKARIN AKITASVSRA GVLSGVGKLV
101  RQGAKFGTRA VPYVGTALLA HDVYETFKED IQARGCRYDP ETDKFVKGYE
151  YANCLWYEDE RRINRTYGCY GVDSSIMRLM PDRSRFPEVK QLMESQMYRL
201  ARPFWNWRKE ELNKLSSLDW NNFVLNRCTF DWNGGGCAVN KGDDFRAGAS
251  FSLGRNPKYK EEMDAKKPEE ILSLKVDADP DKYIEATGYP GYSEKVEVAP
301  GTKVNMGPVT DRNGNPVQVA ATFGRDAQGN TTADVQVIPR PDLTPASAEA
351  PHAQPLPEVS PAENPANNPD PDENPGTRPN PEPDPDLNPD ANPDTDGQPG
401  TSPDSPAVPD RPNGRHRKER KEGEDGGLSC DYFPEILACQ EMGKPSDRMF
451  HDISIPQVTD DKTWSSHNFL PSNGVCPQPK TFHVFGRQYR ASYEPLCVFA
501  EKIRFAVLLA FIIMSAFVVF GSLGGE*
```

After further analysis, the following gonococcal DNA sequence <SEQ ID 275> was identified:

```
  1  ATGGTCACAA AACATACAAA TTTGAATTTT GCGAAATTGT CGATAATTGC
 51  AATTTTGATG ATGTATTCGT TTGAAGCGAA TGCAAATGCA GTAAAAATAT
101  CTGAAACTCT TTCGGTTGAT ACCGGACAAG GCGCGAAAGT TCATAAGTTC
151  GTTCCTAAAT CAAGTAATAT TTATTCATCT GATTTAACAA AAGCGGTAGA
201  TTTAACGCAT ATCCCCACGG GCGCAAAAGC CCGAATCAAC GCCAAAATAA
251  CCGCCAGCGT ATCCCGCGCC GGCGTATTGT CGGGGGTCGG CAAACTTGTC
301  CGCCAAGGCG CGAAATTCGG CACAAGGGCG GTTCCCTATG TCGGAACAGC
351  CCTTTTAGCC CACGACGTAT ACGAAACTTT CAAAGAAGAC ATACAGGCAC
401  GAGGCTGCCG ATACGATCCC GAAACCGACA AATTT
```

This corresponds to the amino acid sequence <SEQ ID 276; ORF72ng-1>:

```
  1  MVTKHTNLNF AKLSIIAILM MYSFEANANA VKISETLSVD TGQGAKVHKF
 51  VPKSSNIYSS DLTKAVDLTH IPTGAKARIN AKITASVSRA GVLSGVGKLV
101  RQGAKFGTRA VPYVGTALLA HDVYETFKED IQARGCRYDP ETDKF
```

ORF72ng-1 and ORF721-1 show 89.7% identity in 145 aa overlap:

```
                      10        20        30        40        50        60
orf72ng-1.pe MVTKHTNLNFAKLSIIAILMMYSFEANANAVKISETLSVDTGQGAKVHKFVPKSSNIYSS
             || |:||||||||||||||||||||||||||||||||:|||||||||:||||||:|: |||
orf72-1      MVIKYTNLNFAKLSIIAILMMYSFEANANAVKISETVSVDTGQGAKIHKFVPKNSKTYSS
                      10        20        30        40        50        60

                      70        80        90       100       110       120
orf72ng-1.pe DLTKAVDLTHIPTGAKARINAKITASVSRAGVLSGVGKLVRQGAKFGTRAVPYVGTALLA
             || |:|||||||||||||||||||||||||||||:||||||:| ||||:||||||||||||||
orf72-1      DLIKTVDLTHIPTGAKARINAKITASVSRAGVLAGVGKLARLGAKFSTRAVPYVGTALLA
                      70        80        90       100       110       120

                     130       140
orf72ng-1.pe HDVYETFKEDIQARGCRYDPETDKF
             |||||||||||||||  :||||||||
orf72-1      HDVYETFKEDIQARGYQYDPETDKFAKVSGX
                     130       140       150
```

Based on this analysis, including the presence of a putative leader sequence and transmembrane domains in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 33

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 277>:

```
  1  ATGAGATTTT TCGGTATCGG TTTTTTGGTG CTGCTGTTTT TGGAGATTAT
 51  GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGCTGG ACGTTGTTTT
101  TGATGGCGGC AGGTTTTGCC GCCGGCGTGC TGATGCTCAG GCAAACCGGG
151  GCTGACCGGT CTTTTATTGG CGGGCGCGGC AATGAGAAGC GGCGGGAAGG
201  TATCCGTTTA TCAGATGTTG TGGCCTATC..
```

This corresponds to the amino acid sequence <SEQ ID 278; ORF73>:

```
  1  MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAAGFA AGVLMLRQTG
 51  LTGLLLAGAA MRSGGKVSVY QMLWPI..
```

Further work revealed the complete nucleotide sequence <SEQ ID 279>:

```
  1  ATGAGATTTT TCGGTATCGG TTTTTTGGTG CTGCTGTTTT TGGAGATTAT
 51  GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGCTGG ACGTTGTTTT
101  TGATGGCGGC AGGTTTTGCC GCCGGCGTGC TGATGCTCAG GCATACGGGG
151  CTGTCCGGTC TTTTATTGGC GGGCGCGGCA ATGAGAAGCG GCGGGAGGGT
201  ATCCGTTTAT CAGATGTTGT GGCCTATCCG TTATACGGTG GCGGCTGTGT
251  GTCTGATGAG TCCGGGATTC GTATCCTCGG TGTTGGCGGT ATTGCTGCTG
301  CTGCCGTTTA AGGGAGGGGC AGTGTTGCAG GCAGGAGGTG CGGAAAATTT
351  TTTCAACATG AACCAATCGG GCAGAAAAGA GGGCTTTTCC CGCGATGACG
401  ATATTATCGA GGGAGAATAT ACGGTTGAAG AGCCTTACGG CGGCAATCGT
451  TCCCGAAACG CCATCGAACA CAAAAAAGAC GAATAA
```

This corresponds to the amino acid sequence <SEQ ID 280; ORF73-1>:

```
  1  MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAAGFA AGVLMLRHTG
 51  LSGLLLAGAA MRSGGRVSVY QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL
101  LPFKGGAVLQ AGGAENFFNM NQSGRKEGFS RDDDIIEGEY TVEEPYGGNR
```

```
151  SRNAIEHKKD E*
```

Computer analysis of this amino acid sequence gave the following results:

Homology with a predicted ORF from *N.meningitidis* (strain A)

ORF73 shows 90.8% identity over a 76aa overlap with an ORF (ORF73a) from strain A of *N. meningitidis:*

```
                    10        20        30        40        50        60
orf73.pep   MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTLFLMAAGFAAGVLMLRQTGLTGLLLAGAA
            |||||||||||||||||||||||||||||||||||| |||||:|||:|||:|||||||
orf73a      MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTLFLMAATFAAGVVMLRHTGLSGLLLAGAA
                    10        20        30        40        50        60

                    70
orf73.pep   MRSGGKVSVYQMLWPI
            |||||:|||| ||| |
orf73a      MRSGGRVSVYXMLWXIRYTVAAVCXMSPGFVSSVXAVLLXLPFKGGAVLQAGGAENFFNM
```

The complete length ORF73a nucleotide sequence <SEQ ID 281> is:

```
  1  ATGAGATTTT TCGGTATCGG TTTTTTGGTG CTGCTGTTTT TGGAGATTAT
 51  GTCGATTGTG TGGGTTGCCG ATTGGTTGGG CGGCGGTTGG ACGCTGTTTC
101  TAATGGCGGC AACCTTTGCC GCCGGCGTGG TGATGCTCAG GCATACGGGG
151  CTGTCCGGTC TTTTATTGGC GGGCGCGGCA ATGAGAAGCG GCGGGAGGGT
201  ATCCGTTTAT CANATGTTGT GGCNTATCCG TTATACGGTG GCGGCGGTGT
251  GTCNGATGAG TCCGGGATTC GTATCCTCGG TGTNGGCGGT ATTGCTGNTG
301  CTNCCGTTTA AGGGAGGTGC AGTGTTGCAG GCAGGAGGTG CGGAAAATTT
351  TTTCAACATG AACCANTCGG GCAGAAAAGA NGGCNTTTCC CGCGATGACG
401  ATATTATCGA GGGGGAATAT ACGGTTGAAG ANCCTTACGG CGGCANTCGT
451  TTCCGAAACG CCNTNGAACA CAAAAAAGAC GAATAA
```

This encodes a protein having amino acid sequence <SEQ ID 282>:

```
  1  MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAATFA AGVVMLRHTG
 51  LSGLLLAGAA MRSGGRVSVY XMLWXIRYTV AAVCXMSPGF VSSVXAVLLX
101  LPFKGGAVLQ AGGAENFFNM NXSGRKXGXS RDDDIIEGEY TVEXPYGGXR
151  FRNAXEHKKD E*
```

ORF73a and ORF73-1 show 91.3% identity in 161 aa overlap

```
                    10        20        30        40        50        60
orf73a.pep  MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTLFLMAATFAAGVVMLRHTGLSGLLLAGAA
            |||||||||||||||||||||||||||||||||||| |||||:|||||||||||||||
orf73-1     MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTLFLMAAGFAAGVLMLRHTGLSGLLLAGAA
                    10        20        30        40        50        60

                    70        80        90       100       110       120
orf73a.pep  MRSGGRVSVYXMLWXIRYTVAAVCXMSPGFVSSVXAVLLXLPFKGGAVLQAGGAENFFNM
            |||||||||| ||| ||||||||| |||||||||| |||| ||||||||||||||||||||
orf73-1     MRSGGRVSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFKGGAVLQAGGAENFFNM
                    70        80        90       100       110       120

                   130       140       150       160
orf73a.pep  NXSGRKXGXSRDDDIIEGEYTVEXPYGGXRFRNAXEHKKDEX
            | |||| | |||||||||||||| |||| | ||| |||||||
orf73-1     NQSGRKEGFSRDDDIIEGEYTVEEPYGGNRSRNAIEHKKDEX
                   130       140       150       160
```

Homology with a predicted ORF from *N.gonorrhoeae*

ORF73 shows 92.1% identity over a 76aa overlap with a predicted ORF (ORF73.ng) from *N. gonorrhoeae:*

```
orf73.pep   MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTLFLMAAGFAAGVLMLRQTGLTGLLLAGAA  60
            |||||||||||||||||||||||||||||||||||| |||||||||:|||:||||||||
```

```
orf73ng      MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTLFLMAATFAAGVLMLRHTGLSGLLLAGAA   60

orf73.pep    MRSGGKVSVYQMLWPI                                               76
             ::|:||||||||||||
orf73ng      VKSSGKVSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFKGGAVLQAGGAENFFNM  120
```

The complete length ORF73ng nucleotide sequence <SEQ ID 283> is:

```
  1  ATGAGATTTT TCGGTATCGG TTTTTTGGTG CTGCTGTTTT TGGAAATTAT
 51  GTCGATTGTG TGGGTTGCCG ATTGGCTGGG CGGCGGTTGG AcgcTGTTTC
101  TAATGGCGGC AACCTTTGCC GCCGGTGTGC TGATGCTCAG GCATAcggGG
151  CTGTCCGGTC TTTTATTGGC TGGCGCGGCG GTAAAAagta gtgGGAAGGT
201  ATCTGTTTAT CagatgtTGT GGCCTATCCG TTATAcggtg gcggcggtgT
251  GTCTGatgag tCcggGATTC GTATCCTccg tgttggCGGT ATTGCTGCTG
301  CTGCcgttta aggGaggGgc agtgttgcag gcaggaggtg cggaaaATTT
351  TTTCAACATg aaCcaatcgg gcagaaAaga gggatttttc cacgatgacg
401  atattatcga gggagaatat acggttgaaa aacctgacgg cggcaatcgt
451  tcccgaAAcg ccatcgaaca cgaaaAagac gaataA
```

This encodes a protein having amino acid sequence <SEQ ID 284>:

```
  1  MRFFGIGFLV LLFLEIMSIV WVADWLGGGW TLFLMAATFA AGVLMLRHTG
 51  LSGLLLAGAA VKSSGKVSVY QMLWPIRYTV AAVCLMSPGF VSSVLAVLLL
101  LPFKGGAVLQ AGGAENFFNM NQSGRKEGFF HDDDIIEGEY TVEKPDGGNR
151  SRNAIEHEKD E*
```

ORF73ng and ORG73-1 show 93.8% identity in 161 aa overlap

```
                    10        20        30        40        50        60
orf73-1.pep  MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTLFLMAAGFAAGVLMLRHTGLSGLLLAGAA
             |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
orf73ng      MRFFGIGFLVLLFLEIMSIVWVADWLGGGWTLFLMAATFAAGVLMLRHTGLSGLLLAGAA
                    10        20        30        40        50        60

                    70        80        90       100       110       120
orf73-1.pep  MRSGGRVSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFKGGAVLQAGGAENFFNM
             ::|:|:||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf73ng      VKSSGKVSVYQMLWPIRYTVAAVCLMSPGFVSSVLAVLLLLPFKGGAVLQAGGAENFFNM
                    70        80        90       100       110       120

                   130       140       150       160
orf73-1.pep  NQSGRKEGFSRDDDIIEGEYTVEEPYGGNRSRNAIEHKKDEX
             ||||||||| :||||||||||||:| |||||||||||:||||
orf73ng      NQSGRKEGFFHDDDIIEGEYTVEKPDGGNRSRNAIEHEKDEX
                   130       140       150       160
```

Based on this analysis, including the presence of a putative leader sequence and putative transmembrane domain in the gonococcal protein, it is predicted that the proteins from *N.meningitidis* and *N.gonorrhoeae*, and their epitopes, could be useful antigens for vaccines or diagnostics, or for raising antibodies.

Example 34

The following partial DNA sequence was identified in *N.meningitidis* <SEQ ID 285>:

```
  1  ATGTTTGTTT TTCAGACGGC ATTCTT.ATG TTTCAGAAAC ATTTGCAGAA
 51  AGCCTCCGAC AGCGTCGTCG GAGGGACATT ATACGTGGTT GCCACGCCCA
101  TCGGCAATTT GGCGGACATT ACCCTGCGCG CTTTGGCGGT ATTGCAAAAG
151  GCG....... .....GCCGA AGACACGCGC GTTACCGCAC AGCTTTTGAG
201  CGCGTACGGC ATTCAGGGCA AACTCGTCAG TGTGCGCGAA CACAACGAAC
251  GGCAGATGGC GGACAAGATT GTCGGCTATC TTTCAGACGG CATGGTTGTG
301  GCACAGGTTT CCGATGCGGG TACGCCGGCC GTGTGCGACC CGGGCGCGAA
351  ACTCGCCCGC CGCGTGCGTG AGGCCGGGTT TAAAGTCGTT CCCGTCGTGG
401  GCGCAAC.GC GGTGATGGCG GCTTTGAGCG TGGCCGGTGT GGAAGGATCC
451  GATTTTTATT TCAACGGTTT TGTACCGCCG AAATCGGGAG AACGCAGGAA
```

```
501  ACTGTTTGCC AAATGGGTGC GGGCGGCGTT TCCTATCGTC ATGTTTGAAA
551  CGCCGCACCG CATCGGTGCA GCGCTTGCCG ATATGGCGGA ACTGTTCCCC
601  GAACGCCGAT TAATGCTGGC GCGCGAAATT ACGAAAACGT TTGAAACGTT
651  CTTAAGCGGC ACGGTTGGGG AAATTCAGAC GGCATTGTCT GCCGACGGCG
701  ACCAATCGCG CGGCGAGATG GTGTTGGTGC TTTATCCGGC GCAGGATGAA
751  AAACACGAAG GCTTGTCCGA GTCCGCGCAA AACATCATGA AAATCCTCAC
801  AGCCGAGCTG CCGACCAAAC AGGCGGCGGA GCTTGCTGCC AAAATCACGG
851  GCGAGGGAAA GAAAGCTTTG TACGAT..
```

This corresponds to the amino acid sequence <SEQ ID 286; ORF75>:

```
  1  MFVFQTAFXM FQKHLQKASD SVVGGTLYVV ATPIGNLADI TLRALAVLQK
 51  A....AEDTR VTAQLLSAYG IQGKLVSVRE HNERQMADKI VGYLSDGMVV
101  AQVSDAGTPA VCDPGAKLAR RVREAGFKVV PVVGAXAVMA ALSVAGVEGS
151  DFYFNGFVPP KSGERRKLFA KWVRAAFPIV MFETPHRIGA ALADMAELFP
201  ERRLMLAREI TKTFETFLSG TVGEIQTALS ADGDQSRGEM VLVLYPAQDE
251  KHEGLSESAQ NIMKILTAEL PTKQAAELAA KITGEGKKAL YD..
```

Further work revealed the complete nucleotide sequence <SEQ ID 287>:

```
  1  ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
 51  ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCTGC
101  GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG
151  CGCGTTACCG CACAGCTTTT GAGCGCGTAC GGCATTCAGG GCAAACTCGT
201  CAGTGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTCGGCT
251  ATCTTTCAGA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG
301  GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGCCGG
351  GTTTAAAGTC GTTCCCGTCG TGGGCGCAAG CGCGGTGATG GCGGCTTTGA
401  GCGTGGCCGG TGTGGAAGGA TCCGATTTTT ATTTCAACGG TTTTGTACCG
451  CCGAAATCGG GAGAACGCAG GAAACTGTTT GCCAAATGGG TGCGGGCGGC
501  GTTTCCTATC GTCATGTTTG AAACGCCGCA CCGCATCGGT GCGACGCTTG
551  CCGATATGGC GGAACTGTTC CCCGAACGCC GATTAATGCT GGCGCGCGAA
601  ATTACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA
651  GACGGCATTG TCTGCCGACG GCAACCAATC GCGCGGCGAG ATGGTGTTGG
701  TGCTTTATCC GGCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG
751  CAAAACATCA TGAAAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC
801  GGAGCTTGCT GCCAAAATCA CGGGCGAGGG AAAGAAAGCT TTGTACGATC
851  TGGCTCTGTC TTGGAAAAAC AAATAG
```

This corresponds to the amino acid sequence <SEQ ID 288; ORF75-1>:

```
  1  MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
 51  RVTAQLLSAY GIQGKLVSVR EHNERQMADK IVGYLSDGMV VAQVSDAGTP
101  AVCDPGAKLA RRVREAGFKV VPVVGASAVM AALSVAGVEG SDFYFNGFVP
151  PKSGERRKLF AKWVRAAFPI VMFETPHRIG ATLADMAELF PERRLMLARE
201  ITKTFETFLS GTVGEIQTAL SADGNQSRGE MVLVLYPAQD EKHEGLSESA
251  QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*
```

Computer analysis of this amino acid sequence gave the following results:

<u>Homology with a predicted ORF from *N.meningitidis* (strain A)</u>

ORF75 shows 95.8% identity over a 283aa overlap with an ORF (ORF75a) from strain A of *N. meningitidis:*

```
                     10        20        30        40        50        60
orf75.pep    MFVFQTAFXMFQKHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKAXXXXXAEDTR
                      ||||||||||||||||||||||||||||||||||||||||||     |||||
orf75a                MFQKHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTR
                              10        20        30        40        50

                     70        80        90       100       110       120
orf75.pep    VTAQLLSAYGIQGKLVSVREHNERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLAR
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf75a       VTAQLLSAYGIQGKLVSVREHNERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLAR
                     60        70        80        90       100       110

                    130       140       150       160       170       180
orf75.pep    RVREAGFKVVPVVGAXAVMAALSVAGVEGSDFYFNGFVPPKSGERRKLFAKWVRAAFPIV
```

```
              ||||:|||||||||| |||||||||| ||||||||||||||||||||||||:|||:|
orf75a        RVREVGFKVVPVVGASAVMAALSVAGVAGSDFYFNGFVPPKSGERRKLFAKWVRVAFPVV
                 120       130       140       150       160       170

                 190       200       210       220       230       240
orf75.pep     MFETPHRIGAALADMAELFPERRLMLAREITKTFETFLSGTVGEIQTALSADGDQSRGEM
              |||||||||:|||||||||||||||||||||||||||||||||||||||:|||:||||||
orf75a        MFETPHRIGATLADMAELFPERRLMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEM
                 180       190       200       210       220       230

                 250       260       270       280       290
orf75.pep     VLVLYPAQDEKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALYD
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
orf75a        VLVLYPAQDEKHEGLSESAQNIMKILTAELPTKQAAELAAKITGEGKKALYDLALSWKNK
                 240       250       260       270       280       290

orf75a        X
```

The complete length ORF75a nucleotide sequence <SEQ ID 289> is:

```
  1  ATGTTTCAGA AACATTTGCA GAAAGCCTCC GACAGCGTCG TCGGAGGGAC
 51  ATTATACGTG GTTGCCACGC CCATCGGCAA TTTGGCGGAC ATTACCCTGC
101  GCGCTTTGGC GGTATTGCAA AAGGCGGACA TCATCTGTGC CGAAGACACG
151  CGCGTTACCG CGCAGCTTTT GAGCGCGTAC GGCATTCAGG GCAAACTCGT
201  CAGCGTGCGC GAACACAACG AACGGCAGAT GGCGGACAAG ATTGTCGGCT
251  ATCTTTCAGA CGGCATGGTT GTGGCACAGG TTTCCGATGC GGGTACGCCG
301  GCCGTGTGCG ACCCGGGCGC GAAACTCGCC CGCCGCGTGC GTGAGGTCGG
351  GTTTAAAGTT GTCCCTGTTG TCGGCGCAAG CGCGGTGATG GCGGCTTTGA
401  GTGTGGCTGG TGTGGCGGGA TCCGATTTTT ATTTCAACGG TTTTGTACCG
451  CCGAAATCGG GCGAACGTAG GAAATTGTTT GCCAAATGGG TGCGGGTGGC
501  GTTTCCCGTC GTGATGTTTG AAACGCCGCA CCGCATCGGG GCGACGCTTG
551  CCGATATGGC GGAACTGTTC CCCGAACGCC GATTAATGCT GGCGCGCGAA
601  ATCACGAAAA CGTTTGAAAC GTTCTTAAGC GGCACGGTTG GGGAAATTCA
651  GACGGCATTG GCGGCGGACG GCAACCAATC GCGCGGCGAG ATGGTGTTGG
701  TGCTTTATCC GGCGCAGGAT GAAAAACACG AAGGCTTGTC CGAGTCCGCG
751  CAAAACATCA TGAAAATCCT CACAGCCGAG CTGCCGACCA AACAGGCGGC
801  GGAGCTTGCC GCCAAAATCA CGGGCGAGGG AAAAAAAGCT TTGTACGATC
851  TGGCACTGTC TTGGAAAAAC AAATGA
```

This encodes a protein having amino acid sequence <SEQ ID 290>:

```
  1  MFQKHLQKAS DSVVGGTLYV VATPIGNLAD ITLRALAVLQ KADIICAEDT
 51  RVTAQLLSAY GIQGKLVSVR EHNERQMADK IVGYLSDGMV VAQVSDAGTP
101  AVCDPGAKLA RRVREVGFKV VPVVGASAVM AALSVAGVAG SDFYFNGFVP
151  PKSGERRKLF AKWVRVAFPV VMFETPHRIG ATLADMAELF PERRLMLARE
201  ITKTFETFLS GTVGEIQTAL AADGNQSRGE MVLVLYPAQD EKHEGLSESA
251  QNIMKILTAE LPTKQAAELA AKITGEGKKA LYDLALSWKN K*
```

ORF75a and ORF75-1 show 98.3% identity in 291 aa overlap:

```
                  10        20        30        40        50        60
orf75a.pep    MFQKHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAY
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
orf75-1       MFQKHLQKASDSVVGGTLYVVATPIGNLADITLRALAVLQKADIICAEDTRVTAQLLSAY
                  10        20        30        40        50        60

                  70        80        90       100       110       120
orf75a.pep    GIQGKLVSVREHNERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREVGFKV
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
orf75-1       GIQGKLVSVREHNERQMADKIVGYLSDGMVVAQVSDAGTPAVCDPGAKLARRVREAGFKV
                  70        80        90       100       110       120

                 130       140       150       160       170       180
orf75a.pep    VPVVGASAVMAALSVAGVAGSDFYFNGFVPPKSGERRKLFAKWVRVAFPVVMFETPHRIG
              |||||||||||||||||| ||||||||||||||||||||||||||:|||:||||||||||
orf75-1       VPVVGASAVMAALSVAGVEGSDFYFNGFVPPKSGERRKLFAKWVRAAFPIVMFETPHRIG
                 130       140       150       160       170       180

                 190       200       210       220       230       240
orf75a.pep m  ATLADMAELFPERRLMLAREITKTFETFLSGTVGEIQTALAADGNQSRGEMVLVLYPAQD
              |||||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
```